

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:05:42 ; Search time 11154 Seconds
(without alignments)
11391.904 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catgtagacggtgccgg.....ttaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3106	100.0	3106	6	AX395268	Sequence
2	3106	100.0	3106	10	AF042822	Mus muscu
3	3102.8	99.9	3248	10	BC005496	Mus muscu
4	2595	83.5	3174	10	AB049189	Rattus no
5	2313.6	74.5	2568	10	AB037898	Rattus no
6	1905.6	61.4	3273	9	BC030532	Homo sapi
7	1889.8	60.8	3149	9	AF118224	Homo sapi
8	1884.8	60.7	3120	9	AF133086	Homo sapi
9	1884.8	60.7	3128	9	AB030036	Homo sapi
10	1883.2	60.6	3147	6	AR081724	Sequence
11	1883.2	60.6	3147	6	AR229704	Sequence
12	1883.2	60.6	3147	6	AR229712	Sequence
13	1883.2	60.6	3147	6	AR229797	Sequence
14	1883.2	60.6	3147	9	AF057145	Homo sapi
15	1883.2	60.6	3147	9	AF081725	Sequence
16	1530.8	49.3	2900	6	AR229705	Sequence
17	1530.8	49.3	2900	6	AR229705	Sequence
18	1530.8	49.3	2900	9	HSU20428	Human SNCL9
19	1175.8	37.9	2148	9	BC005826	Homo sapi
20	945.4	30.4	1823	9	BC018146	Homo sapi
21	809.6	26.1	3487	5	AB038498	Xenopus l
22	701.2	22.6	1553	6	AR263832	Sequence
23	623.2	20.1	225649	2	AC114542	Mus muscu
24	556.4	17.9	726	6	AX473052	Sequence
25	551.8	17.8	723	6	EL3204	Human CDNA
26	359.6	11.6	236387	2	AC096042	Rattus no
27	359.6	11.6	261906	2	AC112632	Rattus no
28	359.6	11.6	277797	2	AC128347	Rattus no
29	306.8	9.9	434	6	AR273737	Sequence
30	306.8	9.9	434	6	AR277318	Sequence
31	306.8	9.9	434	6	AX368770	Sequence
32	265	8.5	591	6	AX261852	Sequence
33	244.8	7.9	3183	10	BC029645	Mus muscu
34	242.4	7.8	3194	9	AY055384	Homo sapi
35	240.8	7.8	2449	9	HSR319876	Homo sapi
36	240.8	7.8	2672	6	AX207905	Sequence
37	240.8	7.8	3104	6	AX207903	Sequence
38	239.2	7.7	1327	6	AX538208	Sequence
39	239.2	7.7	2409	6	AX360098	Sequence
40	239.2	7.7	3143	6	AX538206	Sequence
41	238	7.7	240079	2	AC128364	Rattus no
42	212.2	6.8	174526	9	AP001183	Homo sapi
43	212.2	6.8	182736	2	AC021672	Homo sapi
44	212.2	6.8	190314	9	AC019227	Homo sapi
45	206	6.6	734	6	AX375702	Sequence

ALIGNMENTS

RESULT 1	AX395268	AX395268	3106 bp	DNA	linear	PAT 18-MAY-2002
LOCUS	Sequence 5 from Patent WO0203787.					
DEFINITION	AX395268					
ACCESSION	AX395268					
VERSION	AX395268.1	GI:21066293				
KEYWORDS	Mus musculus (house mouse)					
SOURCE	Mus musculus					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	Allen, K.D. and Leviten, M.W.					
AUTHORS	Transgenic mice containing targeted gene disruptions					
TITLE	Patent: WO 0203787-A 5 17-JAN-2002;					
JOURNAL						

Deltagen, Inc. (US)		Location/Qualifiers		Query Match	
source		1..3106		100.0%; Score 3106; DB 6; Length 3106;	
		/organism="Mus musculus"		Best Local Similarity 100.0%; Pred. No. 0;	
		/mol_type="genomic DNA"		Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
		/db_xref="taxon:10090"			
BASE COUNT		692 a 862 c 897 g 655 t			
ORIGIN					
Qy	901	ATAGCCTGAGCCCATGGAACCCACACGCTGT			

KRSASGVQELKRLRIITHPSFNDFDYDIALLELEKSVEYSTVVRPICLPDATHVPP				
AGKAIWITGWHGTHKEGTGALLIQKGEIRVINOTTCEDLMPQOITPRMCMGVFLSGGV				
DSQCGSGGFLPSAEKDRMFQAGVSWGBGCAQRNKPgyYVTRLPVVRDWIKEHTGV"				
BASE COUNT	692 a	862 c	897 g	655 t
ORIGIN				
Query Match	100.0%; Score 3106; DB 10; Length 3106;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 3106; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	CATGCTAGACGGCTGCCCGAGGACACACGCGTCTGAGACGGGCGATCGAGCCGCCAAAA	60	
Db	1	CATGCTAGACGGCTGCCCGAGGACACACGCGTCTGAGACGGGCGATCGAGCCGCCAAAA	60	
QY	61	CCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAGGACTTCGGCGGGGAC	120	
Db	61	CCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAGGACTTCGGCGGGGAC	120	
QY	121	TCAAGTCAACTCCCGCTAGAGAAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC	180	
Db	121	TCAAGTCAACTCCCGCTAGAGAAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC	180	
QY	181	CTCGGAACAATGCCAAGAAAGTGAGAGAGCGAGGCCCGCAGCGCTGTGGTGGTGTGTFGG	240	
Db	181	CTCGGAACAATGCCAAGAAAGTGAGAGAGCGAGGCCCGCAGCGCTGTGGTGGTGTGTFGG	240	
QY	241	CAGTGTCTGTTTCAGCTTCT	300	
Db	241	CAGTGTCTGTTTCAGCTTCT	300	
QY	301	ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA	360	
Db	301	ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA	360	
QY	361	TCCTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCTCGGCGAGCCAGG	420	
Db	361	TCCTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCTCGGCGAGCCAGG	420	
QY	421	TGAAGAGGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCTGGGTCCCTTACCACAAGA	480	
Db	421	TGAAGAGGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCTGGGTCCCTTACCACAAGA	480	
QY	481	AGTGGGTGTAACTGCGCTTTCAGTGAGGGCAGTGTATCGGCTTACTCTGGTCAAGTTTCA	540	
Db	481	AGTGGGTGTAACTGCGCTTTCAGTGAGGGCAGTGTATCGGCTTACTCTGGTCAAGTTTCA	540	
QY	541	GCATCCCCCAGACCTCGCAGAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA	600	
Db	541	GCATCCCCCAGACCTCGCAGAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA	600	
QY	601	CATTGCCACCCCGAGCACTGAGGAGTGTGAATTCCTTTCGTGTCTAAACATCTGTGTGGCTTCC	660	
Db	601	CATTGCCACCCCGAGCACTGAGGAGTGTGAATTCCTTTCGTGTCTAAACATCTGTGTGGCTTCC	660	
QY	661	CCATTGACCCCGAGTCTGTCAGAGGACTCAGGACAAACAGCTGCAGTTTTCCTCTGCATG	720	
Db	661	CCATTGACCCCGAGTCTGTCAGAGGACTCAGGACAAACAGCTGCAGTTTTCCTCTGCATG	720	
QY	721	CCCATGGTGAGCAGTGACACGCTTCTACTACCCCTGGCTTCCCCAACAGTCCCTTACC	780	
Db	721	CCCATGGTGAGCAGTGACACGCTTCTACTACCCCTGGCTTCCCCAACAGTCCCTTACC	780	
QY	781	CGCATGCCCGCTGCAGTGGGTCTCGGGGGGAGCGCGGACTCTGTGCTGAGCTCACCT	840	
Db	781	CGCATGCCCGCTGCAGTGGGTCTCGGGGGGAGCGCGGACTCTGTGCTGAGCTCACCT	840	
QY	841	TCCGAAGCTTTGATGTGCGTCCCTGTGATGAGCATGGCAGTGTGCTGTGTGTGTGTGTGT	900	
Db	841	TCCGAAGCTTTGATGTGCGTCCCTGTGATGAGCATGGCAGTGTGCTGTGTGTGTGTGTGT	900	
QY	901	ATAGCCTGAGCCCATGGAAACCCACAGCTGTGTGTGGGTGTGTGGCACCTTCTCACCT	960	
Db	901	ATAGCCTGAGCCCATGGAAACCCACAGCTGTGTGTGGGTGTGTGGCACCTTCTCACCT	960	

QY	961	CTTACAACTGACTTTCCTCTCTCTCCAGAACGCTTCTCTTGTTCAGCTGATACCAATA	1020						
Db	961	CTTACAACTGACTTTCCTCTCTCTCCAGAACGCTTCTCTTGTTCAGCTGATACCAATA	1020						
QY	1021	CTCACCGGCGACATCTCTGGCTTTGAGGCCACTTTTCTTCCAGCTGCCCAAGATGAGCAGCT	1080						
Db	1021	CTCACCGGCGACATCTCTGGCTTTGAGGCCACTTTTCTTCCAGCTGCCCAAGATGAGCAGCT	1080						
QY	1081	GTGGCGGCTTTTGTAGTGACACCAAGGAGCATTTAGCAGCCCTTACTATCCAGGCCACT	1140						
Db	1081	GTGGCGGCTTTTGTAGTGACACCAAGGAGCATTTAGCAGCCCTTACTATCCAGGCCACT	1140						
QY	1141	ACCGGCCCAACATCAACTGACATGGAATCAAGGTGCCCAACCAACCGGAACTGGAAG	1200						
Db	1141	ACCGGCCCAACATCAACTGACATGGAATCAAGGTGCCCAACCAACCGGAACTGGAAG	1200						
QY	1201	TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGCTACAGTGGGCTCTGCACCAAGG	1260						
Db	1201	TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGCTACAGTGGGCTCTGCACCAAGG	1260						
QY	1261	ACTATGTGGAGTCAACGGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTGGTAGCA	1320						
Db	1261	ACTATGTGGAGTCAACGGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTGGTAGCA	1320						
QY	1321	GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGACACCGGCT	1380						
Db	1321	GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGACACCGGCT	1380						
QY	1381	TCCTAGTGTAGTACTCTCTTACGACTTCCAAACGACCGCTGCCAGGAGTCTCATGTGCA	1440						
Db	1381	TCCTAGTGTAGTACTCTCTTACGACTTCCAAACGACCGCTGCCAGGAGTCTCATGTGCA	1440						
QY	1441	AGACTGAGCGCTGCATCCGAAAGGAACTGCGCTGCGACGCTGGCAGACTGCCCCGATT	1500						
Db	1441	AGACTGAGCGCTGCATCCGAAAGGAACTGCGCTGCGACGCTGGCAGACTGCCCCGATT	1500						
QY	1501	ATAGTGTAGCGCTTACTGCGGATGCAATGCCACCCACAGTTCAAGTGCACCAACACAGT	1560						
Db	1501	ATAGTGTAGCGCTTACTGCGGATGCAATGCCACCCACAGTTCAAGTGCACCAACACAGT	1560						
QY	1561	TCGTGAGAGCCCTCTTCTGGTCTGTGAAGTGTCAAGTGTGAGGAGCGGAAGTACG	1620						
Db	1561	TCGTGAGAGCCCTCTTCTGGTCTGTGAAGTGTCAAGTGTGAGGAGCGGAAGTACG	1620						
QY	1621	AGGAGGCTGACGCTCTCTGCTGGGAGTTTCAAGTGTTCCTCAATGGGAGTGTCTCCTC	1680						
Db	1621	AGGAGGCTGACGCTCTCTGCTGGGAGTTTCAAGTGTTCCTCAATGGGAGTGTCTCCTC	1680						
QY	1681	AGAGCCAGAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG	1740						
Db	1681	AGAGCCAGAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG	1740						
QY	1741	ACAGCTGGAATGCTCTCTTGCACCAATATACCTACCGCTGCCAAATGGCCTCTGTC	1800						
Db	1741	ACAGCTGGAATGCTCTCTTGCACCAATATACCTACCGCTGCCAAATGGCCTCTGTC	1800						
QY	1801	TGAGCAAGGCGCAACCTCTGAGTGTGATGGGAAGACGAGCTGTAGCGATGGCTCCGATGAGA	1860						
Db	1801	TGAGCAAGGCGCAACCTCTGAGTGTGATGGGAAGACGAGCTGTAGCGATGGCTCCGATGAGA	1860						
QY	1861	AAAACTGTGCTGTGGGCTGCGATCTTTTACCAACAGGCTCGCGTGGTGGTGGACGA	1920						
Db	1861	AAAACTGTGCTGTGGGCTGCGATCTTTTACCAACAGGCTCGCGTGGTGGTGGACGA	1920						
QY	1921	ATGCGACGAGGCGAGTGGCCTTGGCAGTGGCTCCACGCCCTGGGCGAGGCCACT	1980						
Db	1921	ATGCGACGAGGCGAGTGGCCTTGGCAGTGGCTCCACGCCCTGGGCGAGGCCACT	1980						
QY	1981	TGTGTGGGCGCTCGCTCATCTCTCTCTGAGTGGTGTCTCTGAGCTCATTTTCAGG	2040						
Db	1981	TGTGTGGGCGCTCGCTCATCTCTCTCTGAGTGGTGTCTCTGAGCTCATTTTCAGG	2040						

2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGTGACAGGCGCTTCTCGGTCTGCTGG 2100
2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGTGACAGGCGCTTCTCGGTCTGCTGG 2100
2101 ACCAGACAGCGCAGGCTCTGGGTGAGAGCTGAAGCTCAACAGTATCATCAACC 2160
2101 ACCAGACAGCGCAGGCTCTGGGTGAGAGCTGAAGCTCAACAGTATCATCAACC 2160
2161 ACCCTTCTTCAATGATTTCACTTTCGACTATGACATCGCTTCTCGAGCTGGAGAGT 2220
2161 ACCCTTCTTCAATGATTTCACTTTCGACTATGACATCGCTTCTCGAGCTGGAGAGT 2220
2221 CGGTGAGTACAGACCGCTGTGCGGCCATCTCGCTCGCTGATGCTACCATCTCTTCC 2280
2221 CGGTGAGTACAGACCGCTGTGCGGCCATCTCGCTCGCTGATGCTACCATCTCTTCC 2280
2281 CTGCTGCAAGGCCATCTGGTTCACAGCTGGGGGCACACAAAGAGGGAGGTACCGGAG 2340
2281 CTGCTGCAAGGCCATCTGGTTCACAGCTGGGGGCACACAAAGAGGGAGGTACCGGAG 2340
2341 CGCTGATCTCTGCAAGGGTGAGATCGTGTCTATCAACACAGACCATCTGTGAGGACCTCA 2400
2341 CGCTGATCTCTGCAAGGGTGAGATCGTGTCTATCAACACAGACCATCTGTGAGGACCTCA 2400
2401 TGCCGACAGCATCAACCCAGGAATGATGTGTGGGTTTCTCAGTGGGGGTGTGAAT 2460
2401 TGCCGACAGCATCAACCCAGGAATGATGTGTGGGTTTCTCAGTGGGGGTGTGAAT 2460
2461 CTTGCCAGGGTACTCTGGTGGCCCTTGTCAAGCGGGAAGAGTGGCGATGTTC 2520
2461 CTTGCCAGGGTACTCTGGTGGCCCTTGTCAAGCGGGAAGAGTGGCGATGTTC 2520
2521 AGCTGTGTGTGAGTGGGGTGAAGGCTGCGCTCAGAGGAAACAACCCAGGCGGTGTACA 2580
2521 AGCTGTGTGTGAGTGGGGTGAAGGCTGCGCTCAGAGGAAACAACCCAGGCGGTGTACA 2580
2581 CAAGGCTCTCTGTAGTTTGGGACTGGAATCAAGAGCACATCGGGGTATAGCAGCATGGAC 2640
2581 CAAGGCTCTCTGTAGTTTGGGACTGGAATCAAGAGCACATCGGGGTATAGCAGCATGGAC 2640
2641 AGACAGCGGACCAACACACCCAGAGGATGCCGACATGCACACCTGGATACAGAGAG 2700
2641 AGACAGCGGACCAACACACCCAGAGGATGCCGACATGCACACCTGGATACAGAGAG 2700
2701 GAACACTGACGACATTTATCTGTGGCTCCCCCCCCCAACACACACCACTGTGAAT 2760
2701 GAACACTGACGACATTTATCTGTGGCTCCCCCCCCCAACACACCACTGTGAAT 2760
2761 GCATCTTAGACTCAGAGTTCTTCAAGTGGGACCCCTCAAGATTGGAGAGAGAAC 2820
2761 GCATCTTAGACTCAGAGTTCTTCAAGTGGGACCCCTCAAGATTGGAGAGAGAAC 2820
2821 TTGCGTCTAGCGGCCAGCTGTGGGGCAAGGGTTTGTATGCGAGCCTTCCCCCTTAGCC 2880
2821 TTGCGTCTAGCGGCCAGCTGTGGGGCAAGGGTTTGTATGCGAGCCTTCCCCCTTAGCC 2880
2881 CTGAGCTGGTGAAGATGATGCTGTCCGAGAGCTGCTTCCAACTGTCAATTGAGCTCCC 2940
2881 CTGAGCTGGTGAAGATGATGCTGTCCGAGAGCTGCTTCCAACTGTCAATTGAGCTCCC 2940
2941 GGGAGCCCTATGGAGGAGGGCTCAGGCTCACTCTTTTCAAGNAGCGCCAGCCCTAGGA 3000
2941 GGGAGCCCTATGGAGGAGGGCTCAGGCTCACTCTTTTCAAGNAGCGCCAGCCCTAGGA 3000
3001 ACCCGAAGAGAGTGGTACTAAGGCTGAATTTTGTCTGTGTCAGGGGTGGGTAT 3060
3001 ACCCGAAGAGAGTGGTACTAAGGCTGAATTTTGTCTGTGTCAGGGGTGGGTAT 3060
3061 TTGAGAGTAAACATTTATTTCTTTTAAAAAATAAAAAA 3106
3061 TTGAGAGTAAACATTTATTTCTTTTAAAAAATAAAAAA 3106

RESULT 3
BC005496
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BC005496 3248 bp mRNA linear ROD 16-APR-2003
Mus musculus suppression of tumorigenicity 14 (colon carcinoma),
mRNA (CDNA clone MGC:7395 IMAGE:3488059), complete cds.
BC005496
BC005496.1 GI:13529565
MGC.

Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3248)

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquelliano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Madan, A.C., Shrivastava, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, M.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.G., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 3248)

Strausberg, R.

Direct Submission

Submitted (27-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mdpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAC Plate; 7 Row: i Column: 16

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

Location/Qualifiers

1. 3248

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="MGC:7395 IMAGE:3488059"

/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months

old, gross tissue."

/clone_lib="NCI CGAP_Mam5"

/lab_host="DH10B"

FEATURES
source

gene
/note="Vector: pCMV-SPORTc"
1. .3248
/gene="St14"
/note="synonym: Epithin"
/db_xref="LocusID:19143"
/db_xref="MGI:1338881"
173. .2740
/codon_start=1
/product="suppression of tumorigenicity 14 (colon carcinoma)"
/protein_id="AAH05496.1"
/db_xref="GI:13529566"
/db_xref="LocusID:19143"

CDS

/translation="MGSNRRKAGGSGQDFGAGLKYNLSLENNMGFEVEFLPANNA
KKVEKRRPRWVLLVALEFLLSLMAGLLVHHRVNRVQKVFNGHLRTINEIFL
DAYENSTFSTFSLASQVEALKLYNEVPLGSPVHKHSAVAFSEGSVIAYVWSEFS
IPHLAEVDRAVAVRVTLLPPRARALKSVLTSVVAFPIDPRLQRTODNSCSFAL
HAHGAATVTRTPGPNSPVPAHARQWVLRGDADSVLSLTPSPVAPCDSEHSDLV
TVYDLSLSPNEPHAVRLCGTFSPSNLTFSSQVFLVTLINTDRHGFEPATFQL
PMSSCGFLSDTQGFSPYPGHPYNNICTWNIKVNRRNRVRFKFLVLPVNV
PVGSCTKDVEINGEKYGERQFVYSSNSKITYHFDSDHSYTDFTGLAEYLSYDSN
DPCPMFKCTGRCKRKLRCGWADCPYSDERYCRNATHQFTCKNFKPLFWVC
DSVNDCCGDSDEGSCPCGSKNGKCLPOSCKNGKNDGSDGSDASCDVSNVVS
CTKYTYRCNGELCLSKNPECDKTDPCSDGSEKNCDCGLRSFTQARVVGSTNADG
ERPQVSLHALQGHLCGLASLSPDLVLSAAHCFQDDXNFKYSIDYMTWTFGLGLDQS
KRSASGVQLKRLIITHSFNDFTEYDIDALLELEKSVEYSYTVVRPILCPATHVFP
AGKAIWVGWHTKEGTEGALLIQGEIRVINTQCTEDLMPQOITPRMNCVGLSGGV
DSCQDGGPLSSAEXDGMFQAGVYVSWEGCAQRNKPVGVTLPVVRWDIHEHTGV"
BASE COUNT 734 a 918 c 928 g 668 t
ORIGIN

Query Match 99.9%; Score 3102.8; DB 10; Length 3248;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CATGTTAGACGGCTGCGCGGAGGACACACGCTCTGAGACCGCGCATCGACCGCAAAA 60
DB 111 CATGTTAGACGGCTGCGCGGAGGACACACGCTCTGAGACCGCGCATCGACCGCAAAA 170
QY 61 CCATGGTTAGCAATCGGGGCGGCAAGCGCGGAGGGGCTCTCAGGACTTCGGCGGGGAC 120
DB 171 CCATGGTTAGCAATCGGGGCGGCAAGCGCGGAGGGGCTCTCAGGACTTCGGCGGGGAC 230
QY 121 TCAAGTACAACTCCCGCTAGAGAAATGAATGGCTTTGAGGAGGGTGTGGATTCCTGC 180
DB 231 TCAAGTACAACTCCCGCTAGAGAAATGAATGGCTTTGAGGAGGGTGTGGATTCCTGC 290
QY 181 CTGCGAACAAATGCGAAGAGTGGAGAGCGAGCGCCCGCGGCTGGGTGGTGGTGG 240
DB 291 CTGCGAACAAATGCGAAGAGTGGAGAGCGAGCGCCCGCGGCTGGGTGGTGGTGG 350
QY 241 CAGTGTCTTACGTTCTCTTCTCTCCCTCAATGCTGGCTGGCTGGCTGGCTGGCTCC 300
DB 351 CAGTGTCTTACGTTCTCTTCTCTCCCTCAATGCTGGCTGGCTGGCTGGCTGGCTCC 410
QY 301 ATTATCGGAATGCGGGTTCAAAAGTCTTCAATGGCCATCTGAGGATCACAATGAGA 360
DB 411 ATTATCGGAATGCGGGTTCAAAAGTCTTCAATGGCCATCTGAGGATCACAATGAGA 470
QY 361 TCTTTCTGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCTGGCGCAGCAGG 420
DB 471 TCTTTCTGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCTGGCGCAGCAGG 530
QY 421 TGAAGGAGCGCTGAAGCTCTGTCAATGAATGCTGCTGGGTGCTTACCAAGA 480
DB 531 TGAAGGAGCGCTGAAGCTCTGTCAATGAATGCTGCTGGGTGCTTACCAAGA 590
QY 481 AGTCGGTGTAACTGCTTCAAGTGGGAGCTGTATCGCTTACTACTGTGAGATTCA 540
DB 591 AGTCGGTGTAACTGCTTCAAGTGGGAGCTGTATCGCTTACTACTGTGAGATTCA 650
QY 541 GCATCCCCCACACCTGGCAGAGAGAGTGTGATCGCGCATGGCTGTGGAGCGAGTTGTAA 600

DB 651 GCATCCCCCACACCTGGCAGAGAGGTTGATCGCGCATGGCTGTGAGCGAGTTGTAA 710
QY 601 CATTTGCCACCCCGAGCAGCGGCACTGAATCTCTGCTGCTAACTCTGCTGCTGCTCC 660
DB 711 CATTTGCCACCCCGAGCAGCGGCACTGAATCTCTGCTGCTAACTCTGCTGCTGCTCC 770
QY 661 CCATTGACCCCGAGAAATGCTGAGAGGACTCAGGACAAACAGCTGCAGTTTGGCTGCATG 720
DB 771 CCATTGACCCCGAGAAATGCTGAGAGGACTCAGGACAAACAGCTGCAGTTTGGCTGCATG 830
QY 721 CCCATGGTGCAGCAGTGACACCGCTTCACTACCCCTGGCTTCCCCACACAGCTCCCTACCCGG 780
DB 831 CCCATGGTGCAGCAGTGACACCGCTTCACTACCCCTGGCTTCCCCACACAGCTCCCTACCCGG 890
QY 781 CGCATGCCCGCTGCGAGTGGGTCTCTGCGGGGGGAGCGCCGACTCTGTGCTGAGCCTCACCT 840
DB 891 CGCATGCCCGCTGCGAGTGGGTCTCTGCGGGGGGAGCGCCGACTCTGTGCTGAGCCTCACCT 950
QY 841 TCCGAAGCTTTGATGTGCTGCTCCCTGTGATGAGCAGTGGCAGTGCACCTGTCTACCGTGTATG 900
DB 951 TCCGAAGCTTTGATGTGCTGCTCCCTGTGATGAGCAGTGGCAGTGCACCTGTCTACCGTGTATG 1010
QY 901 ATAGCTGAGCCCCCATGGAACCCCAACCGCTGTGCTGCGGCTGTGCGACCTTCTCACCT 960
DB 1011 ATAGCTGAGCCCCCATGGAACCCCAACCGCTGTGCTGCGGCTGTGCGACCTTCTCACCT 1070
QY 961 CCTACAACTGACTTTCTCTCTCTCCCAAGAACGCTCTTCTGCTGAGCTGTATAACCAATA 1020
DB 1071 CCTACAACTGACTTTCTCTCTCTCCCAAGAACGCTCTTCTGCTGAGCTGTATAACCAATA 1130
QY 1021 CTGACCGGCGACATCTGCTGCTTTGAGGCGACTTTCTTCCAGCTGCCCAAGATGAGGAGCT 1080
DB 1131 CTGACCGGCGACATCTGCTGCTTTGAGGCGACTTTCTTCCAGCTGCCCAAGATGAGGAGCT 1190
QY 1081 GTGCGGCTTTTGTGAGTGACACCCCAAGGAGACATTTAGCAGCCCTTACTATCCAGGCACT 1140
DB 1191 GTGCGGCTTTTGTGAGTGACACCCCAAGGAGACATTTAGCAGCCCTTACTATCCAGGCACT 1250
QY 1141 ACCCGCCCAACATCAATGACATGCAATATCAAGGTGCCCCAACACCGGAACGTTGAAGG 1200
DB 1251 ACCCGCCCAACATCAATGACATGCAATATCAAGGTGCCCCAACACCGGAACGTTGAAGG 1310
QY 1201 TCGGCTTCAAATCTTCTATCTGCTGAGACCCCAAGTACAGTGGGCTCTCTGACCAAGG 1260
DB 1311 TCGGCTTCAAATCTTCTATCTGCTGAGACCCCAAGTACAGTGGGCTCTCTGACCAAGG 1370
QY 1261 ACTATGTGAGATCAACGGGGAGAGAGTACTCGGCTGAGAGTCCCAAGTTTGTGTTGAGCA 1320
DB 1371 ACTATGTGAGATCAACGGGGAGAGAGTACTCGGCTGAGAGTCCCAAGTTTGTGTTGAGCA 1430
QY 1321 GCAAACAGCAGCAAGATTACAGTCCATTCTCAATTTCTGATCACTCGTACACGGAACCGGGT 1380
DB 1431 GCAAACAGCAGCAAGATTACAGTCCATTCTCAATTTCTGATCACTCGTACACGGAACCGGGT 1490
QY 1381 TCCTAGCTGAGTACCTCTCTACGACTCCAGACCGCGGCGGAGGAGTTCATGTGCA 1440
DB 1491 TCCTAGCTGAGTACCTCTCTACGACTCCAGACCGCGGCGGAGGAGTTCATGTGCA 1550
QY 1441 AGACTGGACGGTGCATTCGAAAGGAACCTCGCTGCGACCGGCTGGGCGAGACTGCCCGGATT 1500
DB 1551 AGACTGGACGGTGCATTCGAAAGGAACCTCGCTGCGACCGGCTGGGCGAGACTGCCCGGATT 1610
QY 1501 ATAGTGTAGCGGTTACTGTCGCGATGCAATGCAACCAACAGTTTCAAGTTCAGTGCACCAAG 1560
DB 1611 ATAGTGTAGCGGTTACTGTCGCGATGCAATGCAACCAACAGTTTCAAGTTCAGTGCACCAAG 1670
QY 1561 TCTGCAAGCCCTCTCTGCTGCTGAGTGTCAAGTGTTCAGTGTTCAGTGTTCAGTGTTCAGTGT 1620
DB 1671 TCTGCAAGCCCTCTCTGCTGCTGAGTGTTCAGTGTTCAGTGTTCAGTGTTCAGTGTTCAGTGT 1730
QY 1621 AGGAGGGCTGACAGCTGCTCTGCTGCGAGTTTCAAGTGTTCAGTGTTCAGTGTTCAGTGTTC 1680
DB 1731 AGGAGGGCTGACAGCTGCTCTGCTGCGAGTTTCAAGTGTTCAGTGTTCAGTGTTCAGTGTTC 1790

QY	2761	GCATCCTTAGAGACTCAGAGTTCTTCCMAAGTGGGACCCCTCCTCAAGAGTTGGAGAGAGAAC	2820
Db	2871	GCATCCTTAGAGACTCAGAGTTCTTCCMAAGTGGGACCCCTCCTCAAGAGTTGGAGAGAGAAC	2930
QY	2821	TTGCGTGCTAGCGGCCCCAGCGCTGGGGGCAAGGTTTGTATGGCAGCCCTTCCCTCTCTAGCC	2880
Db	2931	TTGCGTGCTAGCGGCCCCAGCGCTGGGGGCAAGGTTTGTATGGCAGCCCTTCCCTCTCTAGCC	2990
QY	2881	CTGAGCTGGGTGAAGATGATGCTGTGTCGGAGAGCTGCTTCCAACTGTCTATTGAGCTCCC	2940
Db	2991	CTGAGCTGGGTGAAGATGATGCTGTGTCGGAGAGCTGCTTCCAACTGTCTATTGAGCTCCC	3050
QY	2941	GGGAGCCCTATGGAGAGAGCGGCTCAGGCTCAGCTCTTTTTCAGGAAGCGCCAGCCCTAGGA	3000
Db	3051	GGGAGCCCTATGGAGAGAGCGGCTCAGGCTCAGCTCTTTTTCAGGAAGCGCCAGCCCTAGGA	3110
QY	3001	ACCCAGAAAGAGTGGTACCTAAGGCTGAAATTTGTTGCTGTTCCAGGGGTGGGTAT	3060
Db	3111	ACCCAGAAAGAGTGGTACCTAAGGCTGAAATTTGTTGCTGTTCCAGGGGTGGGTAT	3170
QY	3061	TTGAGAGTAAACATTTTATTTCTTTTAAAAA	3106
Db	3171	TTGAGAGTAAACATTTTATTTTAAAAA	3216
RESULT	4		
LOCUS	AB049189		
DEFINITION	Rattus norvegicus mRNA for membrane bound arginine specific serine protease, complete cds.		
ACCESSION	AB049189		
VERSION	AB049189.1	GI:10336526	
KEYWORDS	Rattus norvegicus (Norway rat)		
SOURCE	Rattus norvegicus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (sites)		
AUTHORS	Inoue,H., Takahashi,K. and Kishi,K.		
TITLE	membrane-bound arginine specific serine protease		
JOURNAL	Published only in DataBase (2000)		
REFERENCE	2 (bases 1 to 3174)		
AUTHORS	Inoue,H., Takahashi,K. and Kishi,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-SEP-2000) Hideshi Inoue, Tokyo University of Pharmacy and Life Science, School of Life Science, 1432-1 Horinouchi, Hachioji-shi, Tokyo 192-0392, Japan (E-mail:hinoe@ls.toyaku.ac.jp, Tel:81-426-76-7153, Fax:81-426-76-7157)		
FEATURES	Location/Qualifiers		
source	1..3174		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/strain="Wistar"		
	/db_xref="taxon:10116"		
	/sex="male"		
	/tissue_type="duodenum"		
	/dev_stage="7 weeks old"		
	127..2694		
	/codon_start=1		
	/evidence=experimental		
	/product="membrane bound arginine specific serine protease"		
	/protein_id="BAB13765.1"		
	/db_xref="GI:10336527"		
	/translation="MGNNRGRKAGGSDQFGAGKNGRLNMGFEQGVFLPVNNA KOVEXRGPRTFVSLASOVKEALKMYSEVPVLGPYHKSTVTAPEGSVIAYWSEFS DAYNSTSTFVSLASOVKEALKMYSEVPVLGPYHKSTVTAPEGSVIAYWSEFS IIPHLEEVDRAMAVERTVLPPEARAKSLVLTSTVAFIDPRLMQTDNCSGSLV HARGTIVTRTFVTPFNFSGPYPAHARCQWLKRDADSVLSTFRSDPAPCDGSDSLV TYDLSPEHVAVRVLCGTSPSYNLTLFSQNVFLVTLITNDRRPFGEATFPQL PKMSCGGILLSAQGTFSSPYHPYNNITWNIKVPNNRNVKRFKFLVLDVPSN PVGSKTDVLENGEKGFCGERSOFVYVNSSKITVHFDHSYTDTPGLAEVLSYDSN		
CDS			

DPGPMFMCKTRIRKDLRCDGWADCPDYSDEHRCNATHOFMCKNFKPLFWYC
DSVNDGDSDEBSCSPAGSFKNSGKCLPQSQCNKDDCGDSDASCDNNAVS
CTKYTVRCOMGLCLNKNPBCDGGKDCSDGDEKNCDCGLRSFTKQARVVGTTNADBG
EWQVSLHALGQHLGASLISPDWLVSAAHCFODETIFKYSJHMTWTAFLGLDOS
KRSASGVQEHKJLXLIHPHSPNFTFDYDIALELEXPABYSVTVRPICLPDNTHFP
AGKAIWIGHYKEGTGALLQKGEIRVINQTTCELLPQOITPMCMCVGELSGLV
DSQGDGSGPLSSVEKDGRIFFQAGVSWGSCAQRNPKGYTTRPEVRDWIKESQTV"
360
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
513
variation
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
573
variation
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
2121
variation
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
3157.3174
polyA_site 697 a 887 c 911 g 679 t
BASE COUNT
ORIGIN
Query Match 83.5%; Score 2595; DB 10; Length 3174;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 2853; Conservative 0; Mismatches 230; Indels 50; Gaps 7;
QY 1 CATGGTAGCGGTGCGCGGAGGACACACCGTCTGAGACCGGCGATCGGACCGCCAAA 60
DB 65 CATGGTAGTCTGGTTCGCGGAGAGACACACCGTCTGAGACCGGCGATCGGCGCGCGA 124
QY 61 CCATGGGTAGCAATCGGGCGCGAAGCGGCGGCTCTCAGACTTCGGCGCGGAC 120
DB 125 CCATGGGGAACAATCGGGCGCGAAGCGGCGGCTCTCAGACTTCGGCGCGGAC 184
QY 121 TCAAGTAACTCCCGGCTAGAGAACATGAATGCTTTGAGAGGGGTGAGTTCCTGC 180
DB 185 TCAAGTAACTCCCGGCTAGAGAACATGAATGCTTTGAGAGGGGTGAGTTCCTGC 244
QY 181 CTGGGAACAATGCAAGATGAGAGAGGAGCGCCACAGCGCTGGGTGGTGGTGG 240
DB 245 CTGTGAACAACGCGGAGGAGGAGGAGCGCCACAGCGCTGGGTGGTGGTGG 304
QY 241 CAGTGTCTTTCAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
DB 305 CCGTAGTGTTCAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 364
QY 301 ATTATCGGAATGTGCGGGTTCAAAAGTCTTCAATGCGCATCTGAGGATCACAATGAGA 360
DB 365 ATTACCGGAATGTGAGGATCCAAAGTCTTCAATGCGCATCTGAGGATCAAGATGAGA 424
QY 361 TCTTTCTGGATGCGGTATGAGAACCTCACCTCCACAGAGTTTATCAGCTGCGCAGCAG 420
DB 425 ACTTCTGAGCGGTATGAGAACCTCACCTCCACAGAGTTTATCAGCTGCGCAGCAG 484
QY 421 TGAAGAGGCGCTGAAGCTGCTGACAAATGAAGTCCCTGCTGCTGCTGCTGCTGCTG 480
DB 485 TGAAGAGGCGCTGAAGCTGCTGACAAATGAAGTCCCTGCTGCTGCTGCTGCTGCTG 544
QY 481 AGTCCGTGTAACTGCTTCAAGTGGGCGAGTGTATCGCTACTACTGCTCAGAGTTCA 540
DB 545 AATCGATGTAACTGCTTCAAGTGGGCGAGTGTATCGCTACTACTGCTCAGAGTTCA 604
QY 541 GCATCCCCCACCTGCGAGAGAGTGTATCGGCGCATGGCTGTGGAGCGAGTTGTA 600
DB 605 GCATCCCCCACCTGCGAGAGAGTGTATCGGCGCATGGCTGTGGAGCGAGTTGTA 664
QY 601 CATTGCCACCCGAGACCGGACCTGAAATCTCTGCTGCTAACTGCTGCTGCTGCTTCC 660

DB 665 CACTGCCACCCGAGACGCGGCACTGAAATCTCTGCTGCTGCTGCTGCTGCTGCTTCC 724
QY 661 CCATTGACCCCAAGTCTGACAGAGACTCAGACACACAGCTGCAGTTTGGCTGCATG 720
DB 725 CCATTGACCCCAAGTCTCAGAGACTCAGACACACAGCTGCAGTTTGGCTGCATG 784
QY 721 CCATTGATGACAGAGTGAACAGCTTCACTACCCCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 785 CCGTGGCAGGACAGTGAACAGCTTCACTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 844
QY 781 CGATGCGCGCTGCGAGTGGTCTTGGGGGGGAGCGGACTCTGCTGCTGCTGCTGCTGCTG 840
DB 845 CTATGCGCGCTGCGAGTGGTCTTGGGGGGGAGCGGACTCTGCTGCTGCTGCTGCTGCTG 904
QY 841 TCCGAAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 905 TCCGAAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 964
QY 901 ATAGCTGAGCCCATGGAAACCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 965 ACAGCTGAGCCCATGGAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024
QY 961 CTTACAACTGACTTTCTCTCTCCAGAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
DB 1025 CTTACAACTGACTTTCTCTCTCCAGAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1084
QY 1021 CTGACCGGCGACATCTCTGCTTGGAGCCACTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
DB 1085 CTGACCGGCGACATCTCTGCTTGGAGCCACTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1144
QY 1081 GTGGCGCTTTTGTAGTGACACCCAAAGGAGCAATTTAGAGCCCTCTATCTCTCTCTCTCT 1140
DB 1145 GTGGCGCTTTTGTAGTGAGCCAAAGGAGCAATTTAGAGCCCTCTATCTCTCTCTCTCT 1204
QY 1141 ACCGCGCCAACTCAATGCTGACATGCAATATCAAGTGGCCCAACACCGGAACTGAGG 1200
DB 1205 ACCGCGCCAACTCAATGCTGACATGCAATATCAAGTGGCCCAACACCGGAACTGAGG 1264
QY 1201 TGCGCTTCAAACCTCTCTCTCTGCTGAGACCCCAACGCTTACAGTGGGCTCTCTGACCAAG 1260
DB 1265 TACGCTTCAAACCTCTCTCTCTGCTGAGACCCCAACATACAGTGGGAACTGACCAAG 1324
QY 1261 ACTATGTGAGATCAACGCGGAGAACTGCTGCTGCTGAGAGTCCAGTTTGTGGTGAGCA 1320
DB 1325 ACTATGTGAGATCAACGCGGAGAACTGCTGCTGCTGAGAGTCCAGTTTGTGGTGAGCA 1384
QY 1321 GCAACAGCAGCAAGATTTACAGTCCACTTCCATCTGATCACTGCTACAGCAGACCGGGT 1380
DB 1385 GCAACAGCAGCAAGATTTACAGTCCACTTCCATCTGATCACTGCTACAGCAGACCGGGT 1444
QY 1381 TCTAGCTGAGTACCTCTCTCTAGACTCCAAACGACCCCTGCTGCTGCTGCTGCTGCTGCTG 1440
DB 1445 TCTGCTGAGTATCTCTCTCTAGACTCCAAATGACCTTACCTGCTGCTGCTGCTGCTGCTG 1504
QY 1441 AGACTGAGCGGTGATCCGAAAGGAACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
DB 1505 AGACTGAGCGGTGATCCGCAAGGAACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1564
QY 1501 ATAGTATGAGCGGTGATCCGCAAGTGAATGCAATGCAACCCACAGTTTCACTGCAAAAC 1560
DB 1565 ACAGTATGAGCGGTGATCCGCAAGTGAATGCAATGCAACCCACAGTTTCACTGCAAAAC 1624
QY 1561 TCTGCAAGCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
DB 1625 TCTGCAAGCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1684
QY 1621 AGAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
DB 1685 AGAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1744
QY 1681 AGAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
DB 1745 AGAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1804

QY 1741 ACAGCGTAATGTGCTCTCTCTGACAAATATACCTACCGCTGCCAAATGGCCCTCTGTC 1800
Db 1805 ACAATGTGAATGCGCTCTCTGACCAATATATACCTACCGCTGCCAAATGGCCCTCTGTC 1864
QY 1801 TGAGCAAGGGCAACCTCAGTGTGATGGAGAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
Db 1865 TGAACAAGGGCAACCTCAGTGTGATGGAGAGACGGACTGTAGCGATGGCTCCGATGAGA 1924
QY 1861 AAAACTGTGACTGTGGGTGGGATCCTTTTACCAACAGGCTCGCGTGGTGGTGACAGA 1920
Db 1925 AAAACTGTGACTGTGGGTGGGATCCTTTTACCAACAGGCTCGCGTGGTGGTGACAGA 1984
QY 1921 ATGGGACAGAGGGGAGTGGGCCCTGCGAGGTGAGCCTCCACGCCCTGGGGCCAGGSCCACT 1980
Db 1985 ATGGGACAGAGGGGAGTGGGCCCTGCGAGGTGAGCCTCCACGCCCTGGGGCCAGGSCCACT 2044
QY 1981 TGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGACGTCATTTGCTTTTACGG 2040
Db 2045 TGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGACGTCATTTGCTTTTACGG 2104
QY 2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGTGAGCGCCTTCTGCGTCTGCTGG 2100
Db 2105 ATGACAAATTTCAAGTACTCAGACTACAGATGTGAGCGCCTTCTGCGTCTGCTGG 2164
QY 2101 ACCAGAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAAAGCTATCATCAACC 2160
Db 2165 ACCAGAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAAAGCTATCATCAACC 2224
QY 2161 ACCCTCTCTCAATGATTTCACTTCGACTATGATCATCGCCTTGTGAGCTGAGAGAGT 2220
Db 2225 ACCCTCTCTCAATGATTTCACTTCGACTATGATCATCGCCTTGTGAGCTGAGAGAGT 2284
QY 2221 CGGTGGAGTACAGCACCTCTGTCGCCCTCATCTGCTGCTGATGCTACCCATGCTTCC 2280
Db 2285 CGGAGAGTACAGCACTGTGTGGGCCCATCTGCTGCTGATGATGATGATGATGATGAT 2344
QY 2281 CTGTGGCAAGGCCATCTGGGTTCAGGCTCGGGGCAACAAAAGAGAGGAGTACCGGAG 2340
Db 2345 CTGGCGGCAAGGCCATCTGGGTTCAGGCTCGGGGCAACAAAAGAGAGGAGTACCGGAG 2404
QY 2341 CGCTGATCTCAGAGAGGTGATGCTGCTGATCATCAACAGCACACCTGTGAGAGCTCA 2400
Db 2405 CACTGATCTCAGAGAGGAGAGATCGGGGTATCAACAGCACACCTGTGAGAGCTCT 2464
QY 2401 TGGCGCAGCAGATCACCCACCAATGATGTGTGGGTTTCTCAGTGGGGGTGTGACT 2460
Db 2465 TGGCGCAGCAGATCACCCACCAATGATGTGTGGGTTTCTCAGTGGGGGTGTGACT 2524
QY 2461 CTTGCCAGGTGACTCTGGTGGCCCTTTGTCAAGCGCGGAGAAAGATGGCGAATGTTCC 2520
Db 2525 CTTGCCAGGTGATTTCCGGTGGCCCTTTGTCCAGTGTGGAGAAAGATGGCGAATGTTCC 2584
QY 2521 AGGCTGTGTGTGAGCTGGAGTGAAGCTGGCTCAGAGGACAAAGCCAGCGCTGTACA 2580
Db 2585 AGGCTGTGTGTGAGCTGGAGTGAAGCTGGCTCAGAGGACAAAGCCAGCGCTGTACA 2644
QY 2581 CAAGCTCTCTGTAGTTCGGGACTGGATCAAGAGCACACTGGGGTATAGCAGATGAC 2640
Db 2645 CAAGATCCCTGAAGTTCGGGACTGGATCAAGAGCAAACTGGGGTATAGAGCATGGAC 2704
QY 2641 AGACAGCGGACACAAACCCACAGGAGTCCGACATGACACCTCGATACAGGAGAG 2700
Db 2705 AGACACCCACCATGAATCTCGAAAGGGATGCCCGCCACCGGCTACCTGGATACAGCAG 2764
QY 2701 GAACTACAGCAGATTTATGCTGTGGCTCTCCCCCCCCCAACACACCCAGACTGTGACT 2760
Db 2765 GATCGCTGAGGACATTTGTGTGTGGCTCTCCCC-----CACATCCAGAGCTGTGACT 2819
QY 2761 GCATCTCTAGGACTCAGAGTTCTTCCA-----AAGTGGGACCCCTCAAGAGTTGG 2811
Db 2820 GCTTCCCTTAGGACTCAGATCTCTTCCCTGAGTTCCAAAGTGGGACCCCTCAGAGTTGG 2879

QY 2812 AGAGAACTTGCTGTAGCGGCCAGCCTG-----GGGGCAAGGGTTTGATGGCAGCT 2867
Db 2880 CCAGAGATCTTCTGTGTGTGGCCAGCCCGGGTGGGGCAGAGATTGATGGCAGCT 2939
QY 2868 TCCCCCTCTAGCCTTAGCTGGGTGAAGATGATCTGTCCCGGAGAGCTGCTTCC----- 2922
Db 2940 T-CCCTCTTAACCTGAACCTGGTAAAGACGATGCTTCTCTGGAGAACTGCTTCCAGACT 2998
QY 2923 -----AACTCTCATTTGAGCTCCCGGAGCCCTATGGAGGAGGGGCTCAGGCTAC 2973
Db 2999 CTAAGGAAACTATCTGCTGGCTCTCTGGGGCCCTATGGGCAGAGGGGCTGCAGG----- 3054
QY 2974 TCTTTTCAGGAAGCGCCAGCCCTAGGAACCCAGAAAAGAGTGGTACCTAAGGCTGAAT 3033
Db 3055 GTCTTTTCAGGAAGCTCCAGCCCTAGGACCCCCAGAAAAGAGTGGCACCTATGGCTGATT 3114
QY 3034 TGTTTTCTGTGTGCGAGGGTGGGTATTTGAGAGTAAACATTTTATTTCTTTTAAAAA 3093
Db 3115 TGTTTTGTCT-----GGTATTTGAGAGTAAACATTTTATTTCTTTTAAAAA 3161
QY 3094 AAAAAAATAAAAA 3106
Db 3162 AAAAAAATAAAAA 3174

RESULT 5
AB037898
LOCUS AB037898 2568 bp mRNA linear ROD 20-OCT-2001
DEFINITION Rattus norvegicus MBSP mRNA for membrane bound serine protease, complete cds.
ACCESSION AB037898
VERSION AB037898.1 GI:9650963
KEYWORDS membrane bound serine protease; MT-SP1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1
AUTHORS Satomi,S., Yamaaki,Y., Tsuzuki,S., Hitomi,Y., Iwanaga,T. and Fushiki,T.
TITLE A role for membrane-type serine protease (MT-SP1) in intestinal epithelial turnover
JOURNAL Biochem. Biophys. Res. Commun. 287 (4), 995-1002 (2001)
MEDLINE 21458307
PUBMED 11573963
REFERENCE 2 (bases 1 to 2568)
AUTHORS Tsuzuki,S.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Satoshi Tsuzuki, Kyoto University, Graduate School of Agriculture, Division of Food Science and Biotechnology, Laboratory of Nutrition Chemistry; Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail:tkmone@kais.kais.kyoto-u.ac.jp, Tel:81-75-753-6263, Fax:81-75-753-6264)
FEATURES
source Location/Qualifiers
1..2568
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Wistar"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="jejunum"
/dev_stage="8 weeks old"
1..2568
/gene="MBSP"
1..2568
/gene="MBSP"
/codon_start=1
/product="membrane bound serine protease"
/protein_id="BAB03502.1"
/db_xref="GI:9650964"
/translation="MGNNRGRKAGGSGQDFGAGIKYNSRLNMMNPFEGVEFLPVNNA"

Db 1861 GCGGACGAGGCGAGTGCGCCCTTGGAGGTGAGCCTCCACGCTCTGGGCGAGGCGACCTTG 1920
Qy 1983 TGTGGGGCTCGCTCATCTCTCTGACCTGGCTGGCTGCTGAGCTCATCTGCTTTCAGGAT 2042
Db 1921 TGTGGGGCTCGCTCATCTCTCTGACCTGGCTGGCTGCTGAGCTCATCTGCTTTCAGGAT 1980
Qy 2043 GACAAAAATTTCAAGTACTCAGACTACACGATGTGGAGCGGCTTCTGGGTCTGCTGGAC 2102
Db 1981 GAGACAAATTTCAAGTACTCAGACACACACCATGTGGAGCGCTTCTGGGTCTGCTGGAC 2040
Qy 2103 CAGAGCAAGCGAGTGGCTCTGGGGTGAGAGCTGAAGCTCAAGCTATCATCAACCAC 2162
Db 2041 CAGAGCAAGCGAGTGGCTCTGGGGTGAGAGCTGAAGCTCAAGCTATCATCAACTCAC 2100
Qy 2163 CTTCTCTTCAATGATTTCACTTTCAGCTATGACATCGCTTGTGCTGAGCTGAGAGTCG 2222
Db 2101 CTTCTCTTAAATGACTTTTCACTTTCAGCTATGACATCGCTTGTGCTGAGCTGAGAGTCG 2160
Qy 2223 GTGGAGTACAGACCGCTGCTGCGCCCATCTGCTGCTGATGCTTACCATGTCTTCCCT 2282
Db 2161 GCAGAGTACAGACCTGCTGCGCCCATCTGCTGCTGATGATATACCATGTCTTTCCT 2220
Qy 2283 GTGGCAAGGCCATCTGGGTACAGCTGCGGGGACACAAAGAGGAGGTACCGGAGC 2342
Db 2221 GCCGCAAGGCATCTGGGTACAGCTGCGGGGACACAAAGAGGAGGTACCGGAGC 2280
Qy 2343 CTGATCTCTGCAAGAGGTGAGATCGCTGCTCAACACAGACACCTGTGAGACCTCATG 2402
Db 2281 CTGATCTCTGCAAGAGGAGATCGGGTCAACACAGACACCTGTGAGAACTCTTG 2340
Qy 2403 CGCAGCAGATACACCCACGAAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACTCC 2462
Db 2341 CGCAGCAGATACACCCACGAAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACTCC 2400
Qy 2463 TGCAGGCTGACTCTGGTGGCCCTTGTCAAGCGGGAAGAGATGGCGAATGTCAG 2522
Db 2401 TGCAGGCTGATTCGGTGGCCCTTGTCCAGTGGGAAGAGATGGCGAATCTTCAG 2460
Qy 2523 GCTGGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCGGCTGTACACA 2582
Db 2461 GCTGGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCGGCTGTACACA 2520
Qy 2583 AGGCTCCTGTAGTTCGGGACTGGATCAAGAGACACATCGGGGTATAG 2630
Db 2521 AGGATCCTGAAGTTCGGGACTGGATCAAGAGCAAACTGGGGTATAG 2568

RESULT 6
BC030532
LOCUS BC030532 3273 bp mRNA linear PRI 20-MAY-2002
DEFINITION Homo sapiens, suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin), clone MGC:40392 IMAGE:5213189, mRNA, complete cds.
ACCESSION BC030532
VERSION BC030532.1 GI:20988874
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3273)
Direct Submission
Strausberg, R.
Submitted (07-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hri.nih.gov
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Laric, J., Legaspi, R., Maduro, Q.L., Masiello, C., Maki, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Teurigne, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAK Plate: 64 Row: k Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES
source

1..3273
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:40392 IMAGE:5213189"
/tissue_type="Blood, adult leukocytes"
/clone_lib="NIH_MGC_118"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
118..2685
/codon_start=1
/product="suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin)"
/protein_id="AAH30532.1"
/db_xref="GI:20988875"
/translation="MGSDRARKGGGPKDFGAGLKYNSRHKVNGLEEGVEFLPVNVV KYVEKHPGHWVLAALVILGLIIVLGVPLVHLOYDRVQVQVFNMYRITNENFV DAVNSNSTFVSLAKSKVDALILYSGVPLGPKYHESAVTAPESSGATAYKSEFS IPQHLVEARVMAERVMVLPARRSLKSFVTSVAVPTDSTVQRTQNSCSFGL HARGVELMRFPTPGPDPSPYPAHARQWALRGDADSVLSLTITFRFDLASCDEGDLV TVNTLSPEMFHALVLCGTYPSPYNTLTHSSQVLLITITNTERHPGEATFFQL PRMSSCGRLKAKQGTENSPYGHYPDNTWNIENPVNKHVRFKFLLEPGV PAGTCPKDYVEINGEKYQSVFVTSNSKLTIVRSHDSQVTDCTFLAFLYSDSS DPCPGOFTGTCIRKELRCDWADCTDHSDELNCSCDAGHOFCKNPKFLFWYC DSVNDCGNSDEQCSCPAQTFRCNSGKLSKSCQCKNGKDCGDSGSEASCPRVNVIT CTXHTYRLNGLCLSKNPECDGKDCGSDGSDXDCGLRSFTRQARVGTDADEG EWPQVSLHALGQGHICGASLISPNLWLSAAHCYIDDRFYSPTQWVTAFLGLHDS ORGAPGVQERRLKRIISHPPFNDFDYDIALLEKPAYSVMVRPCLPDASHVFP AGKAIWTVGHTGYGCTGALILKQGEIRVINTOTTENLLPOQITPRMVCVPLSGGV DSCQDSSGGLSSVEADGRIFQAGVSVSGDCAQRNKPQVYTRLPFLFDMWKENTGV"

CDS

BASE COUNT 690 a 1001 c 993 g 589 t
ORIGIN
Query Match 61.4%; Score 1905.6; DB 9; Length 3273;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 2260; Conservative 0; Mismatches 514; Indels 6; Gaps 4;
Qy 1 CATGGTAGACGGCTGCCCGGAGGACACACGGCTGTGAGACGGCGGATCGGACCCGCAAAA 60
Db 57 CATGGAGCGCGCGCGCGGAGGACACACGGCTGTGAGACGGCGGCGGCGGCGGCGG 115
Qy 61 CCATGGGTAGCAATCGGGCGGCGGAGGCGGCGGCTCTCAGGACTTCGGCGGGGAC 120
Db 116 CCATGGGAGGAGTTCGGGCGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 175
Qy 121 TCAAGTACAACTCCCGGCTAGAGAACTAATGCTTTTGAGAGGGGTGTGGAGTTCCTGC 180
Db 176 TCAAGTACAACTCCCGGCGGAGGAGGAGTGAATGCTTTTGAGAGGAGGCTGGAGTTCCTGC 235
Qy 181 CTGCGAACTATGCAAGAAAGTGTGAGAGGAGGCGGCGGCGGCGGCGGCGGCGG 240


```
QY 2401 TGCCGACGACATCACCCACGAATGATGTGTGGGTTTCTCAGTGGGGGTGTGGACT 2460
Db 2456 TGCCGACGACATCACCGCGCATGATGTGTGGGTTTCTCAGCGCGCGTGTGGACT 2515
QY 2461 CCTGCGAGGTGACTCTGTGTGGCCCTTGTCAAGCGCGGAGAAAGATGGGGAATGTTC 2520
Db 2516 CCTGCGAGGTGATTCGCGGGGACCCCTGTCAGCGTGGAGCGGATGGCGGATCTTTC 2575
QY 2521 AGGCTGGTGTGTGAGCTGGGCTGAAGCTGGCTCAGAGGAACAAGCCAGCGGTGTACA 2580
Db 2576 AGGCGGTGTGTGAGCTGGGAGACGGCTGGCTCAGAGGAACAAGCCAGCGGTGTACA 2635
QY 2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGGGTATAGCAGCATGCAC 2640
Db 2636 CAAGGCTCCCTGTGTTTCGGGACTGGATCAAGAGAAACACTGGGGTATAGGGGCGGG-- 2693
QY 2641 AGACAGCGGACCAACAAACACCCACAGGATGCCGACATGCACACCTGGATACAGGAGAG 2700
Db 2694 -GCCACCAAAATGTGTACACCTGCGGGGCCACCCATGTCACCCCAAGTGTGCACG-CCT 2751
QY 2701 GAACACTCAGCAGATTTATGTGTGGCTCCCGCTCCCGCCCAACACACACCCAGACTGTGA 2760
Db 2752 CGAGCTGGAGACTGGACCGCTGACTGCACGAGCGCC-CCAGAACATACACTGTGA 2810
QY 2761 GCATCCTTAGGACTCAGAGT 2780
Db 2811 CAATCTCCAGGCTCCAAAT 2830

RESULT 7
AF118224 3149 bp mRNA linear PRI 30-DEC-1999
DEFINITION Homo sapiens matriptase mRNA, complete cds.
ACCESSION AF118224
VERSION AF118224.2 GI:6647301
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2955)
AUTHORS Lin, C.Y., Anders, J., Johnson, M., Sang, Q.A. and Dickson, R.B.
TITLE Molecular cloning of cDNA for matriptase, a matrix-degrading serine
proteinase with trypsin-like activity
J. Biol. Chem. 274 (26), 18231-18236 (1999)
JOURNAL 99303581
MEDLINE 10373424
PUBMED
REFERENCE 2 (bases 1 to 2955)
AUTHORS Lin, C.Y., Anders, J., Johnson, M., Sang, Q.A. and Dickson, R.B.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1999) Lombardi Cancer Center, Georgetown
University Medical Center, 3970 Reservoir Road NW, Washington, DC
20007, USA
REFERENCE 3 (bases 1 to 3149)
AUTHORS Oberst, M.D., Dickson, R.B. and Lin, C.Y.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Lombardi Cancer Center, Georgetown
University Medical Center, 3970 Reservoir Road NW, Washington, DC
20007, USA
REMARK Sequence update by submitter
COMMENT On Dec 30, 1999 this sequence version replaced gi:5359674.
FEATURES
source
1. .3149
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
36. .2603
/function="degrade extracellular matrix"
/note="serine protease with trypsin-like activity;
contains Clr/s and LDL receptor domains"
/codon_start=1
/product="matriptase"
```

```
/protein_id="AAD42765.2"
/db_xref="GI:6647302"
/translation="MGSDRAKKGKGGPKDFAGLKYNSRHKVNGLEKGVLEFVFNV
KKKXGPGFGRVWVLAALIGLLVLLGFLVHLOVDRVQVFNKFMIRITNFV
DAYENSTSEVSLASKVKCALKLLYSGVPFLGPHKESAVTASEGVIATYWFSEF
IPEHLVEARVMAEVRVWMLPPRARSLKSPVTSVAFPTDSTKVTQNSCSFGL
HARVNTLSPMEPHALVQLGCTVPPSPYPAHARCOWALRGDADSLSTRFSLDASCBERSDLV
TVYNTLSPMEPHALVQLGCTVPPSPYPAHARCOWALRGDADSLSTRFSLDASCBERSDLV
PRMSSCGRLKAKGCTNSPYPGHYPPNIDCTWNIEVFNQKVPKFFLLEPGV
PAGTCXDYVEINGEKYKERSQFVTSNSNKITVRHSDQSYTDGFLAEYLSYDSS
DPCPGQTCRTGRICRKLKCDGWADCTDHSDELNCSDCAGHQTCKNFKCPKLVFVC
DSVNDGDNDSQCCPAOTFCRNGKCLSKSQCKNGKDDCGSDGSEASCPKYNVT
CTHYTECLNGCLCLSKNPECDEKDCSDGDEKDCGLRSTFRQARVVGCTDADG
EWPQVSLHALGQGHICGASLISPNWLSAAHCYIDRGRFYSPTQWTAFLGHDQS
QRSPGVQERLRKLIISHPFNDFTFDIDALLELEKPAEYSWVRPCLDPAHSHVFP
AKAIWYGMHTQYGGTGALILQKBEIRVINOTTCENLLPQOITPRMCMCVGFSGGV
DSCQDGGSGPLSVSEADGRIFQAQVSWGDCQAQRNKPVVYTRFLPLFRDMIKENTGV"
BASE COUNT 647 a 959 c 962 g 581 t
ORIGIN
Query Match 60.8%; Score 1889.8; DB 9; Length 3149;
Best Local Similarity 81.4%; Pred. No. 0;
Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;
QY 50 GACCGCCAAACCACTGGTAGCAATCGGGGCCGCAAGCGCGAGGGGCTCTCAGGACTT 109
Db 23 GGCCTCGGGGACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCCCGAGAGACTT 82
QY 110 CGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAAACATGAATCGCTTTGAGGAGGTGT 169
Db 83 CGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAAAGTGAATGGCTTGGAGGAGCGT 142
QY 170 GGAGTCTCTGCTCGGACAAATCCAGAAAGTGGAGAGCGAGGCGGCGGCTGGGT 229
Db 143 GGAGTCTCTGCTCGGACAAACGTCAGAAAGTGGAAAGCATGGCGGGCGGCTGGGT 202
QY 230 GGTGCTGTGGGAGTGTTCAGTCTCTCTGCTCTCCCTCATGCTGGTGTGCTGTGCT 289
Db 203 GGTGCTGTGGGAGCGTGTGATCGGCTCTCTCTGCTGTGGGATCGGCTTCTGCT 262
QY 290 GTGCACTTCCATTCGGAATGTGGGTTCAAAAAGTCTTCAATGGCCACTGAGGAT 349
Db 263 GTGCACTTTCAGTACCGGAGCGTGGCTGCCAGAGTCTTCAATGGCTACATGAGAT 322
QY 350 CACAAATGAGATCTTTCGATGGATGGTATGAGACTCCACCTCCACAGAGTTATCAGCT 409
Db 323 CACAAATGAGAAATTTTGTGATGCTCAGAGAACTCCCACTCAGTTTGAAGCT 382
QY 410 GGCAGCGCAGGTGAAGGAGCGGCTGAAAGTGTCTGTACAATGAAGTCCCTGTCTGGGTCC 469
Db 383 GGCAGCAAGGTGAAGGAGCGGCTGAAAGTGTCTGTACAAGGAGTCCCATTCCTGGGCC 442
QY 470 CTACCAAGAAGTCTGGCTGTAACTGCTTCAAGTGGGAGTGTCTATCGCTACTACTG 529
Db 443 CTACCAAGAAGTCTGGCTGTGAGCGCTTTCAGCGGAGGAGGCTCATCGCTACTACTG 502
QY 530 GTCAAGTTCAGCATCCCGCCACACCTGGCAGAGAGGTTGATCGCGCCATGGCTGGA 589
Db 503 GTCTGAGTTTCCCGCAGACCTCCGAGAGTGTGGAGGAGGCGGAGGCTCATGGCGGAGGA 562
QY 590 GCGAGTTGTAACATGTCACCCCGAGCACCGGCACTGAAATTCCTTCGTGTGCTAATCTGT 649
Db 563 GCGGCTAGTTCATGTCGCCCGCGGGCGCTCCCTGGAAGTCTCTTGTGTGCTCAGT 622
QY 650 GGTGGCTTCCCATTTGACCCCGAGATGCTGAGAGGACTCAGGACAAACAGCTGCACTT 709
Db 623 GGTGGCTTTCCTCCCGAGCTCCAAAACAGTACAGAGGAGCCAGGACAAACAGCTGCACTT 682
QY 710 TGCCCTGTCATGCGCATGTGTCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCAACAG 769
Db 683 TGCCCTGTCAGCGCGCGGTGTGAGCTGTGCTTCAACACCGCGGCTTCCCTGACAG 742
QY 770 TCCCTACCCCGCGCATGCGCGCTGCCAGTGGGTCTCTGGGGGGGAGCGCGACTCTGTGCT 829
```

Db 743 CCCCACCCCGCTCATGCCCGCTGCCAGTGGCGCTCGCGGGGAGCGCCGACTCAGTGTCT 802
QY 830 GAGCCTCACCTTCGGAAGCTTTGATGTCGCTCCCTGTGTGATGAGATGGCAAGTCACTTGGT 889
Db 803 GAGCCTCACCTTCGGAAGCTTTGATGTCGCTCCCTGTGTGATGAGATGGCAAGTCACTTGGT 862
QY 890 CAGCGTGTATGATAGCTCGAGCCCATGGAACCCACGCTGTGTGTGGCTGTGTGGCAC 949
Db 863 GACGGTGTACAAACACCTGAGCCCCATGGAGCCCCACGCTGTGTGGCAC 922
QY 950 CTTCTCACCTCTTACAACTGATTTTCCTCTCTCTCCAGAACGTCCTTGTGTGACGCT 1009
Db 923 CTACCTCTCTCTTACAACTGATTTTCCTCTCTCTCCAGAACGTCCTTGTGTGACACT 982
QY 1010 GATAACCAATACAGCGCGGACATCTGTGCTTTGAGGCCACATTTCTTCCAGCTGCCAA 1069
Db 983 GATAACCAACACTGAGCGCGGACATCTGTGCTTTGAGGCCACATTTCTTCCAGCTGTCTAG 1042
QY 1070 GATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCAGGAGACATTTAGCAGCCCTACTA 1129
Db 1043 GATGAGCAGCTGTGGAGCGCTTTACGTAAAGCCAGGAGACATTTCAACAGCCCTACTA 1102
QY 1130 TCCAGGCCACTACCCGCCCAACATCAACTGCAATGCAATGATGAGTGCCTCAACACCG 1189
Db 1103 CCGAGGCCACTACCCGCCCAACATTTGACTGCATGGAACATTTGAGGTGCGCCAAACCA 1162
QY 1190 GAACTGAGGTGGCTTCAAACTTTCTTATCTGTGGACCCCAACGCTTACAGTGGGCTC 1249
Db 1163 GCATGTGAGGTGGCTTCAAACTTTCTTATCTGTGGACCCCGGCTGTCTGGGGCAC 1222
QY 1250 CTGACCAAGGACTATGTGAGATCAACCGGGAGAGTACTGGGTGAGAGTCCCAGTT 1309
Db 1223 CTGCCCCAAGGACTATGTGAGATCAATGGGAGAAATCTGGGAGAGAGTCCCAGTT 1282
QY 1310 TGTGTTGAGCAGCAACAGCAGAGATTAAGTCCACTTCTTCAATCTGATCACTGATACAC 1369
Db 1283 CGTGTGACCAAGCAACAGCAGAGATTAAGTCCACTTCTTCAATCTGATCACTGATACAC 1342
QY 1370 GGACACGGGTTCTAGCTGAGTACCTCTCTTACGACTCCCAACGACCCGTCGCCAGGAT 1429
Db 1343 CGACACGGGTTCTAGCTGAATACCTCTCTTACGACTCCCAACGACCCGTCGCCAGGAT 1402
QY 1430 GTTCTATGCAAGACTGGAAGGCTGCAATCGAAAGGAACTCGCTGCGACGCGTGGGCGA 1489
Db 1403 GTTCTATGCAAGACTGGAAGGCTGCAATCGAAAGGAACTCGCTGCGACGCGTGGGCGA 1462
QY 1490 CTGCGCGGATTTATGATGATGAGCGTTACTTCCCGATGCAATGCCACCCAGTTCAGCTG 1549
Db 1463 CTGACCGACCAAGCGATGAGCTCAACTGCTGCTGCGACCGCGGCAACAGTTCAGCTG 1522
QY 1550 CAAAAACCACTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGA 1609
Db 1523 CAAGAACAACTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGA 1582
QY 1610 CGGAAGTGAAGAGGCTGAGCTGTCTCTGTGGAGTTTCAAGTTCCTCAATGGAA 1669
Db 1583 CAACAGCGACAGCAGGCTGAGCTGTCTGCGGCGGACCTTCAAGTGTTCCTCAATGGAA 1642
QY 1670 GTGCTCTCCCTCAGACGACAGAGTGAATGGGAGGACCACTGTGGAGATGGGTCTGACGA 1729
Db 1643 GTGCTCTCGAAAGCCAGCTGCAATGGGAGGAGCACTGTGGGAGCGGGTCCGACGA 1702
QY 1730 GGCTTCAATGTGACAGCGTGAATGTCTCTTTCACCAAAATATACCTTACCGCTGCCAAA 1789
Db 1703 GGCTCTCTGCCCCAAGGTGAACGTCTGCTCACTTGTACCAAAACACCTTACCGCTGCTCAA 1762
QY 1790 TGGCTCTGTCTGAGCAAGGCGCAACCTGTAGTGTGATGGGAAGCAGCTGTAGCAGATGG 1849
Db 1763 TGGCTCTGTCTGAGCAAGGCGCAACCTGTAGTGTGATGGGAAGCAGCTGTAGCAGATGG 1822
QY 1850 CTCGATGAGAAACCTGACTGTGGGCTGCGATCTTTTACCAACAGGCTCGGTGTT 1909

Db 1823 CTCAGATGAGAGGACTGCGACTGTGGGCTGGGTCAATTCAGAGACAGGCTCGTGTGT 1882
QY 1910 TGGTGCAAGATGCGGACGAGGCGAGTGGCCCTTGGCAGGTGAGCTTCCACCCCTGGG 1969
Db 1883 TGGGGGCAAGATGCGGATGAGGCGAGTGGCCCTTGGCAGGTGAGCTTCCATGCTCTGGG 1942
QY 1970 CAGGGGCCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGCTGCTCTCTGACGCTCA 2029
Db 1943 CAGGGGCAACATCTGCGGTCTTCCCTCATCTCTCCCACTGGCTGTCTCTCGCGACA 2002
QY 2030 TTGCTTTTCAGGATGACAAAATTTTCAAGTACTCAGACTACACGATGTGGACGCGCTTCT 2089
Db 2003 CTGCTATATGATGACAGAGGATTCAGTACTCAGACCCACGACGAGTGGACGCGCTTCT 2062
QY 2090 GGGTCTGTGGACGAGCAGCGGAGTGGCTTGGGGTGGAGGCTGAGCTGAACTCAACG 2149
Db 2063 GGGCTTTCAGCAGCAGCAGCGCGCCCTTGGGGTGGAGGCTGAGGCTGAGGCTCAAGG 2122
QY 2150 TATCATCACCCACCTTCTTCAATGATTTTCACTTTCGACTATGACATCGCTTGTGGA 2209
Db 2123 CATCATCTCCACCTTCTTCAATGACTTCACTTTCGACTATGACATCGCTTGTGGA 2182
QY 2210 GCTGAGAGTGTGGTGGATGACAGCACCTGTGTGGCCCACTCTGCTGCTGATGCTAC 2269
Db 2183 GCTGAGAGAACCGCAGAGTACAGCTTCCATGTGTGGCGCCATCTGCTGCGGACGCTC 2242
QY 2270 CCATGTCTTCCCTGTGGCAAGGCCATCTGGGTTCAGGGCTGGGGACACCCAGTATGG 2329
Db 2243 CCATGTCTTCCCTGTGGCAAGGCCATCTGGGTTCAGGGCTGGGGACACCCAGTATGG 2302
QY 2330 AGTACCCGAGCGCTGATCTCTGAGAGGCTGAGATCCGTGTGTCATCAACAGACCACTG 2389
Db 2303 AGGCACTGGCGCTGATCTCTCAAAAGGCTGAGATCCGCTCATCAACAGACCACTG 2362
QY 2390 TGAGGACCTATGCGCGCAGAGATCAACCCCAAGATGATGTGTGGGTTCCTCAGTGG 2449
Db 2363 CGAGAACCTCTCTGCGCAGCAGATCAACCGCGCATGATGTGGTGGGTTCCTCAGCG 2422
QY 2450 GGGTGTGACTCTCTGCGAGGCTGACTCTGTGGCCCTTGTCAAGCGCGGAGAAAGATGG 2509
Db 2423 CGGCTGTGACTCTCTGCGAGGCTGATTTCCGGGGGACCCCTGTCCAGGTGAGCGGATGG 2482
QY 2510 GCGAATGTTCCAGGCTGTGTGTGAGCTGGGTGAAAGCTGCGCTCAGAGGAAACAAGCC 2569
Db 2483 GCGGATTTCCAGGCGGTGTGTGAGCTGGGAGACGCTGCGCTCAGAGGAAACAAGCC 2542
QY 2570 AGCGGTGACAGAGGCTCCCTGTGTGTGAGTGGAGTCAAGAGACACACTGGGGTATA 2629
Db 2543 AGCGGTGACAGAGGCTCCCTGTGTGTGAGTGGATCAAGAGAAACAACAAGGATATA 2602
QY 2630 GCAGCATGACAGACAGCGGACCAACAAACACACAGGATGCCGACATGCACACTGG 2689
Db 2603 GGGGCGGG---GCCACCAATGTGTACCTGTGGGGGACCCCATCTCCACCCAGT 2659
QY 2690 ATACAGGAGGAAACATGACGACATTTATGTGTGGCTTCCCCCCCCCAACACACCA 2749
Db 2660 GTGCAGC-CCTGCGGCTGAGACTGGACCGCTGACTGTGACCAAGCGGCC-CCAGACATA 2717
QY 2750 GACTGTGAACCTGATCTTTAGGACTCAGAT 2780
Db 2718 CACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2748

RESULT 8

AF133086

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AF133086 3120 bp mRNA linear PRI 01-OCT-1999
Homo sapiens membrane-type serine protease 1 mRNA, complete cds.
AF133086
AF133086.1 GI:6002713

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 3120)
 Takeuchi, T., Shuman, M.A. and Craik, C.S.
 Reverse biochemistry: use of macromolecular protease inhibitors to
 dissect complex biological processes and identify a membrane-type
 serine protease in epithelial cancer and normal tissue
 Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11054-11061 (1999)
 99432178
 MEDLINE
 PUBMED
 10500122
 2 (bases 1 to 3120)
 Takeuchi, T., Shuman, M.A. and Craik, C.S.
 Direct Submission
 Submitted (04-MAR-1999) Dept. Pharm. Chem., University of
 California, San Francisco, 513 Parnassus Ave., Box 0446, San
 Francisco, CA 94143-0446, USA
 Location/Qualifiers
 1. 3120
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_line="PC-3"
 /cell_type="prostate carcinoma"
 /note="MT-SPI; chymotrypsin clan serine protease"
 /codon_start=1
 /product="membrane-type serine protease 1"
 /protein_id="AAF00109.1"
 /db_xref="GI:6002714"
 /translation="MGSDRARKGGGPKDFAGIKYNSRHEKVNLEEGVEFLPVNV
 KVEKSGPGRWVLAALVLLGLVGLFVWHLQYRDVVRKQVNGYMRIINENPV
 DAVNSSTFVSLAAKDALLLYSGVFLGYPYHESAVTAFSGSVIAYNVSFS
 IPIHLVEAEERVAERVMPLPRASLKFVTVSVAFPTDSKTQRTQDNCSEFL
 HARGVELMRTTFPGLPDPSPAPARQWALRGDADSLVLSLFRSLDASDERGSDLV
 TVNTLSMPBPHALVOLCGYPPSYNLTFFHSQNVLLITLTNTERRHGFGEATFPL
 PRMSSCGRLRKAGQFNSPYHPYDNDCTWNIENPNQHVKVPFPPFILLRPGV
 PACTCPDYIEINGEKYGBRSQFVTSNKNITVRFHDSQSTDTGFLAELYSVDS
 DPCGQCTGIRKRLKCDWADCTHSDLNCSDDAGHQTCKNFKCFKFLFVC
 DSVNDCDSDGSCPAQTFRCNSKCLSKQCNQDGDGSDDEASCPKVNVT
 CTHYRCLNGLGSKNPECDKEDCDGSLRSTFQARVVGVDADBG
 EWPQVSLHAGOHCHISPLNDPMDYDIALELEKPAEYSSWVRPILCLPDASHVFP
 ORSAPGVQERLARIISGFENFTPDYDIALELEKPAEYSSWVRPILCLPDASHVFP
 AGKALVWGHTGYGGLTALLIQKEIRVINGTTCENLLPQITPRMCMCVGLSGV
 DSCQSGSLPSVADGRIFQGVSWGSGCAQRNKPQVYTRPLPLFRDWIKENTGV"
 BASE COUNT 629 a 956 g 571 t
 ORIGIN
 Query Match 60.7%; Score 1884.8; DB 9; Length 3120;
 Best Local Similarity 82.8%; Pred. No. 0;
 Matches 2164; Conservative 0; Mismatches 447; Indels 1; Gaps 1;
 34 CTGAGACCGCGCATCGACCGCCAAACCATGGGTAGCAATCGGGGCGCGCAAGCGCGAG 93
 9 CTGAGACCGCGAGCGG-CCTCGGGGACCATGGGAGCGATCGGGCGCGCAAGCGCGAG 67
 94 GGGGCTCTCAGGACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAAATGAATG 153
 68 GGGGCGCGAAGGACTTCGGCGGGGACTCAAGTACAACTCCCGGCGCGAAGAAAGTGAATG 127
 154 GCTTTCAGGAGGGGTGGAGTTCCTGCTCGCAACAAATGCAAGAAAGTGGAGAGCGAG 213
 128 GCTTTCAGGAGGGGTGGAGTTCCTGCTCGCAACAAATGCAAGAAAGTGGAGAGCGAG 187
 214 GCCCGAGCGCTGGGTGGTCTGGTGGAGTTCCTGCTCGCAACAAATGCAAGAAAGTGGAG 273
 188 GCCCGAGCGCTGGGTGGTCTGGTGGAGTTCCTGCTCGCAACAAATGCAAGAAAGTGGAG 247
 274 TGCTGCTGCTGGGTGGTCTGGTGGAGTTCCTGCTCGCAACAAATGCAAGAAAGTGGAG 333
 248 GGATCGGCTTCCTGGGTGGGATTTTCAGTACCGGAGCGTGGTGGTGGAGAGGCTTCA 307
 334 ATGGCCATCTGAGGATCACAAATGAGATCTTCTGGATGCTATGAGAACTCCACCTCCA 393
 308 ATGGCTACATGAGGATCACAAATGAGATTTTGGATGCTATGAGAACTCCACCTCCA 367

QY 394 CAGAGTTTATCAGCTCGCCAGCCAGAGGAGGCGCTGAAGCTGCTGTACAAATGAAG 453
 Db 368 CTGAGTTTGAAGCCTTGGCCAGAGAGTGAAGGAGCGCTGAAGCTGCTGTACAGCGAG 427
 QY 454 TCCCTGTCCTGGGTCCTTACCAACAAGAGTGGGTGAATGCTGCTTCAAGTGAAGGCGAGT 513
 Db 428 TCCCAITTCCTGGGCCCTTACCAACAAGAGTGGGTGAATGCTGCTTCAAGTGAAGGCGAG 487
 QY 514 TCATCGCTTACTACTGCTAGAGTTTACAGATTCCTCCCAACCTTGGCAGAGAGAGTGAATC 573
 Db 488 TCATCGCTTACTACTGCTAGAGTTTACAGATTCCTCCCAACCTTGGCAGAGAGAGTGAATC 547
 QY 574 GCGCATGCTGCTGGAGCGAGTTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 633
 Db 548 GCGTATGCGCGGAGGAGCGGTAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 607
 QY 634 TCGTGTACATCTGTGGTGGCTTCCCAATTCACCCAGAGTTCGTCAGAGGAGTCAAG 693
 Db 608 TTGTGTTCACCTCAGTGGTGGCTTCCCAACGAGTTCACCAACAGTACAGAGGAGTCAAG 667
 QY 694 ACAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
 Db 668 ACAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
 QY 754 CTGGCTTCCCAACAGTTCCTTACCCGCGCATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 813
 Db 728 CCGGCTTCCCTGACAGCCCTTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
 QY 814 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 873
 Db 788 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
 QY 874 ATGGCAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 933
 Db 848 GCGGCGAGCGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
 QY 934 TGGGCT 993
 Db 908 TGGAGTTTGTGGCAGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
 QY 994 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1053
 Db 968 TCTGCT 1027
 QY 1054 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1113
 Db 1028 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1087
 QY 1114 TTAGCAGCGCTTACTATCCAGGCGCTTACCCGCGCAACATCACTGCAATGCAATGAATCA 1173
 Db 1088 TCAACAGCGCTTACTATCCAGGCGCTTACCCGCGCAACATGCAATGCAATGAATCA 1147
 QY 1174 AGTTCGCGCAACACCGGAGCGTGAAGTGGCTTCAAACTTCTATCTGCTGCTGCTGCTGCTGCT 1233
 Db 1148 AGTTCGCGCAACACCGGAGCGTGAAGTGGCTTCAAACTTCTATCTGCTGCTGCTGCTGCTGCT 1207
 QY 1234 AGTTCAGTGGGCTTCTGCAACAGGAGTATGCTGAGATCAACGCGGAGAGTACTGCG 1293
 Db 1208 GGTGCTTCTGGGCGCTTCTGCGCGCGCTGCTGCGCGCGCTTACGTTGAAGTCAATGGGAGAA 1267
 QY 1294 GTGAGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1353
 Db 1268 GAGAGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1327
 QY 1354 CTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1413
 Db 1328 CAGATCAGTCTTACACCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1387
 QY 1414 ACCCGTGGCGGAGTTCATGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1473
 Db 1388 ACCCATGCTGGCGGAGTTCATGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447

```
1474 GCGACGCTGGCAGACTGCCGGAATATATAGTGATGAGCGTTACTGCGGATGCAATGCCA 1533
1448 GTGATGGCTGGCGGACTGTCACCGACACACGATGAGTCAACTGAGTTGCGACGCG 1507
1534 CCACACAGTTACGTCGCAAAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTCAGTG 1593
1508 GCCACCAAGTTACGTCGCAAAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTCAGTG 1567
1594 TCAACGACTGTGGGACGGAAGTGCAGGAGGCGCTCAGCTGTCTGTGGGAGTTTCA 1653
1568 TGAACGACTGCGGAGACAACAGCAGCAGCAGGCTGATGTTCCGCGCCACGACTTCA 1627
1654 AGTGTTCATAGGAAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGCAACTGTG 1713
1628 GGTGTTCATAGGAAGTGTCTCCGAAAGCCAGCAGTGCATGGGAAGCAGCTGTG 1687
1714 GAGATGGTCTGACGAGGCTTCATGTACAGCGTGAATGTCTCTTTGACCAATATTA 1773
1688 GGGACGGGTCCGACGAGGCTCTCTGCCCAAGGTGAACGTCGTCACTGTACCAACACA 1747
1774 CCTACCGCTGCCAAATGGCTCTGTCTGACGAGGCAACCTCAGTGTGATGGGAAGA 1833
1748 CTTACCGCTGCCCAATGGCTCTGTCTGACGAGGCAACCTCAGTGTGACGGAGG 1807
1834 CGACTGTAGCGAGTGTCTCCGATGAGAAACCTGTGACTGTGGCTGCGATCCTTACCA 1893
1808 AGGACTGTAGCGAGTGTCTCAGTGAAGGACTGCGACTGTGGCTGCGGTCACTACGA 1867
1894 AACAGGCTCCGCTGTGTGGGACGAAATCGCAGGAGGCGAGTGCCTGCGAGTGA 1953
1868 GACAGGCTCTGTGTGGGGACGAGTCCGATGAGGCGGAGTGCCTTGGCAGGTAA 1927
1954 GCTTCCAGCCCTGGGCCAGGCGCACTGTGTGGGCGCTCGCTCATCTCTCTGACTGGC 2013
1928 GCCTGCATGTCTGGGCCAGGCGCACATCTGCGGTGCTTCCCTCATCTCTCCCACTGGC 1987
2014 TGGTCTCTGAGCTCATCTGTTTTCAGATGACAAAATTTCAAGTACTCAGATACAGA 2073
1988 TGGTCTCTGCGGACACTGTCTATCATGATCAGAGGATTCAGGTACTCAGACCCACGC 2047
2074 TGTGGACGCGCTTCTCGGTCTGTGACACGAGCAGCGCAGTGCCTCTGGGCTGAGG 2133
2048 AGTGGACGCGCTTCTGGGCTGTGACACGACGAGCAGCGCCCTGGGCTGAGG 2107
2134 AGTGAAGCTCAACGCTATCATCACCCACCTTCTTCAATGATTTTCACTTCGACTATG 2193
2108 AGCGAGGCTCAAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTCGACTATG 2167
2194 ACATCGCTTGTGGAGCTGGAGAGTCCGTGGAGTACAGACCGTGTGCGGCCCATCT 2253
2168 ACATCGGCTGTGGAGCTGGAGAAACCGGACAGATACAGTCCATGTTGCGGCCCATCT 2227
2254 GCCTGCTGATGTCTACCATGTCTTCTCCTGCTGGCAGGCGCATCTGGGTCAAGGCTGGG 2313
2228 GCCTGCGGAGCGCTCCCATGTCTTCTCCTGCGGACAGGCGCATCTGGGTCAAGGCTGGG 2287
2314 GGCACACAAAGAGGAGTACCGGAGCGCTGATCTCTGACAGAGGAGTACATCCGTGTCA 2373
2288 GACACACCCAGTATGAGGACCTGGCGCGCTGATCTCTGCAAAAGGCTGAGATCCGCGTCA 2347
2374 TCAACAGACACCTGTGAGGACCTCATGCGGACGAGATCAACCCACGATATGTGTG 2433
2348 TCAACAGACACCTGTGAGGAACTCTCTGCGGACGAGATCAACCCGCGCATGATGTGG 2407
2434 TGGGTTTCTCAGTGGGGGTGTGACTCTCTCCAGGCTGACTCTGTGGGCCCTTGTCAA 2493
2408 TGGGTTTCTCAGCGCGGCTGGAATCTCTGCGAGGATGATTCGGGGGACCCCTGTCCA 2467
2494 GCGGGAAGAGTGGGCGGAATGTTCCAGCTGTTGAGCTGGGTGAGAGGCTGGG 2553
2468 GCCTGAGGCGGATGGGCGGATCTTCCAGCGCGTGTGTGAGTGGGAGACGCTGCG 2527
2554 CTCAGAGAACAGCCAGGCGGTGTACACAGGCTCTCTGTAGTTCGGGACTGGATCAAG 2613
```

```
2528 CTCAGAGGACAGCCAGCGGTGTACACAGGCTCCCTCTGTTCCGAGCTGATCAAAG 2587
2614 AGCACACTGGGGTATAGCAGCATGAGACACA 2645
2588 AGAACACTGGGGTATAGGGGCCCGGGGCACCCA 2619

RESULT 9
AB030036 3128 bp mRNA linear PRI 06-JAN-2001
LOCUS Homo sapiens mRNA for prostamin, complete cds.
DEFINITION AB030036
ACCESSION AB030036
VERSION 1 GI:12249014
KEYWORDS prostamin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Yamaguchi, N. and Mitsui, S.
TITLE Molecular cloning of a novel transmembrane serine protease
expressed in human prostate
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3128)
AUTHORS Yamaguchi, N. and Mitsui, S.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1999) Nozomi Yamaguchi, Kyoto Prefectural
University of Medicine, Res. Ins. Geriatrics; Kawaramachi Hirokoji,
Kyoto, Kyoto 602-8566, Japan (E-mail:nozomi@koto.kpu-m.ac.jp,
Tel:81-75-251-5848, Fax:81-75-251-5848)
FEATURES
Location/Qualifiers
1..3128
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="prostate"
16..2583
/codon_start=1
/product="prostamin"
/protein_id="BAB20376.1"
/db_xref="GI:12249015"
/translation="MGSDRARKGGGPKDFGAGLKYSRHEKVNGLGEGVEFLPVNVV
KXVEHGRGVRWVLAALVILIGLIGLVHLYQYRDVVRQKVFNGYMRITNFV
ISQVNSFEVSLASKVDALKLYSGVPLGPKHSAVAFSEGVSVIAYWSFES
IQHLEVEAERVMARERVMLPPARSLKSFVTSVVAPTQSKVQRPODNCSEGL
HARGVELRPTTGPFPDSEYPHARCOALRGDADSVLSLTFSEDLASCDERGSLLV
TYNTILSNRPHALVOLGCTYPPSTNLTPHSSQNVLLITLITERRHPGFATFPQL
PMSSCGGRLKRAQGTFFNSPPYHPNIDCTWNIENQHVKVRFFFLBEPV
PAGTCPKDYVEINGEKYCGERSQFVTSNSKITVRFHSDQSYDTDTGLAEYLSYDS
DPCQGTCTGTCIRKELKELCDGWADCTHSDLCNCDAGHQFTCKNFKPLFWVC
DSVNDGDNSSDEQSCPAOTPCRSNGKCLSKSQCNKDCGDCGSDASCPKVVVT
CTKHYRCNLGLCLSKGNPCDCKEDSCDCKDCGLRSTQARVVGCTDADG
EWPQVSLHALQGHICGASLISPNWLYSAHCYIDDRGFYSDDTQWTVFLGLHDQS
QKSAFGVRLKRIISHPFNDFTFDIDALLELEKPAEYSSMRPCLIPDASHVFP
AKGLIWTGHTQYGGTGLILQREIVRINQTTCEMLLPQOITPRMVCVGLSGV
DSCQSDSGPLSSVEADGRIFQAGVVGWDGCAQRNKPQVYTRLPFRDWIKENTGV"
BASE COUNT 637 a 954 c 956 g 581 t
ORIGIN

Query Match 60.7%; Score 1884.6; DB 9; Length 3128;
Best Local Similarity 81.4%; Pred. No. 0;
Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;

QY 50 GACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAGACTT 109
Db 3 GGCCTCGGGGACATGGGAGGATCGGGCCCGCAAGGGCGGAGGGGCCCAAGACTT 62
QY 110 CGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGAGAGGCT 169
Db 63 CGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGAGAGGCT 122
QY 170 GGAGTTCCTGCTCGAACAATGCCAAGAAAGTGAGAGAGCGAGGCCCGCGCTGGGT 229
```

Db 123 GAGTTCTTCCAGTCAACAACGTCAGAAAGTGAAAGCATGGCCCGGGCGCTGGT 182
QY 230 GGTCTGTGTGAGTGTGTTGAGTCTTCTGCTCTCTCCATCATGGCTGGCTGCTGGT 289
Db 183 GGTCTGTGAGCGGTGCTGATCGGCTCTCTGCTGCTTGTGCTGGGATCGGCTTCTGGT 242
QY 290 GTGGCACTTCCATTATCGGAATGTGGGTTTCAAAAAGTCTTCAATGGCCATCTGAGAT 349
Db 243 GTGGCATTTGAGTACCGGGACGTGCTGTCAGAAAGTCTTCAATGGCTACATGAGAT 302
QY 350 CACAAATGAGATCTTCTGATGCTGTATGAACTTCCACCTCCACAGAGTTTATCAGCT 409
Db 303 CACAAATGAGATTTTGTGATGCTTACGGAACCTTCCAACTCCATGAGTTTGTAAAGCT 362
QY 410 GGCCAGCCAGGTGAAGGAGCGCTGAAGCTGCTGTACAAATGAAGTCCCTGCTGGGTCC 469
Db 363 GGCCAGAGGTGAAGGAGCGCTGAGCTGCTGTACAGCGAGTCCCATCTCTGGGCCC 422
QY 470 CTACCAAGAGTCCGGCTGTAACTGCTTCACTGAGGGCAGTGTATCGCCTACTACTG 529
Db 423 CTACCAAGAGTCCGGCTGTGAAGCTTTCAGCGAGGGCAGCTCATCGCTACTACTG 482
QY 530 GTCAGAGTTGAGATCCCGCCACACTCTGCGAGAGAGTGTATCGCGCATGCGCTGTGGA 589
Db 483 GTCTGAGTTCAAGTATCCCGCAGCACTGTGTGAGGAGGCGCGTCACTGGCCGAGGA 542
QY 590 GCGAGTTGTAACTATGCCACCCCGAGCAGCGGCACTGAAATCTTCTGCTGCTAAATCTGT 649
Db 543 GCGGTGTCTGTCTGCTCCCGCGGCGGCTCTCTGAGTCTCTTGTGCTCACTCAGT 602
QY 650 GGTGCTTCCCTTACCCATGACCCCGAATGCTCGAGAGGACTCAGGACACAGCTGAGTTT 709
Db 603 GGTGCTTTCCTCCAGCACTCCAAACAGTACAGAGGACCCAGAGCAACAGCTGAGCTT 662
QY 710 TGCCCTGATCGCATGTGAGCAGTACACGCTTCACTACCTCTGCTTCCCGCCACAG 769
Db 663 TGGCTGACGCGCGGTGTGAGTGTATGCTGCTTCAACCGCCCGCTTCCCTGACAG 722
QY 770 TCCCTACCGCGCATGCCGCTGCGAGTGGTCTCTGCGGGGAGCGCGACTGTGCT 829
Db 723 CCCCTACCCGCTCATGCCGCTGCACTGGGCTCTGCGGGGAGCGCGACTGCTGCT 782
QY 830 GAGCTCACTTCCGAGCTTGTATGCTGCTCTCTGATGAGCATGCGAGTCACTGCT 889
Db 783 GAGCTCACTTCCGAGCTTGTACCTTGTGCTGCTGCGAGCGCGGCGAGCTGCT 842
QY 890 CACGCTGTATGATGAGCTGAGCCCATGGAACCCACCGCTGTGTGCGGCTGTGTGGAC 949
Db 843 GACGCTGTACACACCTGAGCCCATGAGCCCATGAGCCCATGAGCTGTGTGGAC 902
QY 950 CTCTTCACTTCAACCTGACTTCTCTCTCCAGAACTCTCTTCTTGTGACGCT 1009
Db 903 CTACCTTCCCTTACAACTGACTTCTCTCTCCAGAACTCTCTTCTGCTCATCACT 962
QY 1010 GATTAACCAATGATGAGCGGAGATCTGCTGCTTGTAGGCCACTTCTTCCAGTGTCCAA 1069
Db 963 GATTAACCAATGATGAGCGGAGATCTGCTGCTTGTAGGCCACTTCTTCCAGTGTCCAA 1022
QY 1070 GATGAGCAGCTGTGGGCTTTTGTGATGAGTACCCAGGAGCATTTAGCAGCCCTACTA 1129
Db 1023 GATGAGCAGCTGTGGGCTTTTGTGATGAGTACCCAGGAGCATTTCAACAGCCCTACTA 1082
QY 1130 TCCAGGCACTACCCCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAACACCG 1189
Db 1083 CCCAGGCACTACCCCAACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1142
QY 1190 GAACTGAGGTGCTTCAACTTCTTCTATCTGTGAGCCCGCCACAGTACAGTGGGCTC 1249
Db 1143 GATGTGAGGTGCTTCAAACTTCTTCTATCTGTGAGCCCGCCAGTGTGCTGCGGGCAC 1202
QY 1250 CTGACACAGGACTGTGAGATCAACGGGAGAGTACTTGGCTGAGAGGTCCAGTT 1309

Db 1203 CTGCCCCAAGGACTAGCTGGAGATCAACGGGAGAAATACTCGGAGAGAGTCCAGTT 1262
QY 1310 TGTGTTGAGAGCAACAGCAGCAAGATTAAGTCCACTTTCATTTCTGATCTACTCTGACAC 1369
Db 1263 GGTGTGATCCAGCAACAGATCAAGATTCAGTTTCTGTTCCACTCAGATCAGTCTACAC 1322
QY 1370 GGACACCGGGTCTTAGCTGAGTACTCTCTAGACTTCCAAAGACCCGCTGCCCCAGGAT 1429
Db 1323 CGACACCGGCTTCTTAGCTGAATACCTTCTCTAGCTCCAGTACCCATGATGCCGGGCA 1382
QY 1430 GTTCATGTGAAAGCTGGAACGGTGCATCCGAAAGAACTGCTGCTGCGAGCTGGGCGAGA 1489
Db 1383 GTTCAGCTGCGGCAACGGGCGGTGTATCCGAAAGAGCTGCTGCTGATGCTGGGCGGA 1442
QY 1490 CTGCCCCGATTTATGATGAGCTTACTGCGGATGCAATGCCACCCACCACTTACCTG 1549
Db 1443 CTGACCGACACAGCGATGAGTCAACTGAGTTGCGACCGCGGCCACCACTTACCGT 1502
QY 1550 CAAAAACCAAGTCTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGA 1609
Db 1503 CAAGAACAAAGTTCTGCAAGCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGGGAGA 1562
QY 1610 CGGAAGTGAACAGAGGCTGCACTGCTCTGCTGGGAGTTTCAAGTGTCCAAATGGAA 1669
Db 1563 CAACAGCGACAGCAGGGGTGCACTTGTCCGCGCCAGACCTTCAAGTGTCCAAATGGAA 1622
QY 1670 GTGTCTCCCTCAGAGCCAGAGTCTTAATGGAAAGCACTGTGGAGTGGTCTGACGA 1729
Db 1623 GTGCTCTCGAAGACCAAGTGTCAATGGAAAGCACTGTGGGAGCGGTCCGACGA 1682
QY 1730 GCTTCTATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGCTGCCAAA 1789
Db 1683 GGCCTCTGCCCCAAGTGAACGCTGCTCTTGTACCAACACACCTTACCGCTGCTCAA 1742
QY 1790 TGGCTCTGTGAGAGGAGGCAACCTCTGAGTGTATGGGAAGACGCACTGTAGCGATG 1849
Db 1743 TGGCTCTGTGAGAGGAGGCAACCTCTGAGTGTATGGGAAGGAGCACTGTAGCGAG 1802
QY 1850 CTCGATGAGAGAAATGTGACTGTGGGCTGCGATCTTTTACCAACAGGCTCGCGTGT 1909
Db 1803 CTCGATGAGAGAAATGTGACTGTGGGCTGCGATCTTTTACCAACAGGCTCGCGTGT 1862
QY 1910 TGTGCGCAACAAATCGGACAGGCGAGTGGCCCTTGGCAGGTGAGCTTCCACGCGCTGG 1969
Db 1863 TGGGCGCACGATCGGATGAGGCGAGTGGCCCTTGGCAGGTGAGCTTCCAGTCTGCG 1922
QY 1970 CCAGGCGCACTGTGTGGGCTCGCTCACTCTCTGAGCTGGCTGTGCTCTGAGCTCA 2029
Db 1923 CCAGGCGCACTGTGTGGGCTCGCTCACTCTCTGAGCTGGCTGTGCTCTGAGCTCA 1982
QY 2030 TTGCTTTTCAAGTACCAAAATTTCAAGTACTCAGACTACACGATGTGGAGCGCTTCT 2089
Db 1983 CTGCTATCATGATGACAGAGATTCAGGTACTCAGACCCACGAGTGGACGCTTCTCT 2042
QY 2090 GGTGTCTGTGGAACAGAGCAAGCGAGTCCCTTGGGCTGAGGAGTCAAGCTCAACG 2149
Db 2043 GGTGTGTGCAACAGAGCAAGCGAGCGCCCTTGGGCTGAGGAGCGAGCTCAAGCG 2102
QY 2150 TATCATCAACCCCTTCTTCAATGATTTCACTCTGATATGACATCGCTTGTGGA 2209
Db 2103 CATCATCTCCACCCCTTCTTCAATGATTTCACTCTGATATGACATCGCTTGTGGA 2162
QY 2210 GCTGGAAGTGTGGTGTGAGTACAGACCGCTGCTGCGGCCCATCTGCTGTGATGTAC 2269
Db 2163 GCTGGAAGTGTGGTGTGAGTACAGACCGCTGCTGCGGCCCATCTGCTGTGATGTAC 2222
QY 2270 CCATGTCTTCCCTGTGGCAAGCCATCTGTGGTCAAGGCTGGGGGCAACAAAGAGG 2329
Db 2223 CCATGTCTTCCCTGTGGCAAGCCATCTGTGGTCAAGGCTGGGGGCAACAAAGAGG 2282
QY 2330 AGGTACCGGAGCGCTGATCTTGCAGAGGAGTGTGATCGGTGTATCAACCGAGCACCTG 2389
Db 2283 AGGCACTGGGCGCTGATCTCTGCAAAAGGAGTGTGATCGGCTCATCAACAGACCGCTG 2342

QY 1425 GGGATGTTTCATGTCAGACTGCGAGCGGTGCAATCCGAAAGGAACATGCGCTGCGACGGCTGG 1484
DB 1385 GGGCAGTTTCAGTCCGCGACCGGGCGGTGTATCCGAAAGAGCTGCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCGGATATTATGATGATGAGCGTTACTGCCGATGCAATGCCACCCACAGATTC 1544
DB 1445 GCCGACTGCAACCGACACAGCGATGAGTCAACTGCAGTTGCGACGCCGGCCACAGATTC 1504
QY 1545 ACGTGCAAAACACAGTTCTCTGCAAGCCCTCTCTTGGGTCTGTGACAGTGTCAAGGACTGT 1604
DB 1505 ACGTGCAAGAACAGTTCTCTGCAAGCCCTCTCTTGGGTCTGTGACAGTGTGAAAGACTTGC 1564
QY 1605 GGGAGCGGAAGTGCAGAGAGGGCTGACAGTCTCTCTGCGGAGTTTCAAGTGTTCGAAT 1664
DB 1565 GGAGACAACAGCGACGAGCGAGCGGTGCAGTTGTTCGGGCCAGACCTTCAGGTGTTCGAAT 1624
QY 1665 GGGAGGTGTCTCCCTCAGAGCGACGAGTGTATGGGAAGGACAACTGTGGAGATGGGTCT 1724
DB 1625 GGGAGGTGCCCTCTCGAAAGCCAGCAGTGCATGGGAAGGACGACTGTGGGAGCGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGCAACAAATATACCTACCGCTGC 1784
DB 1685 GACGAGGCTCTCTGCCCAAGGTGAACTGTCTCTCTGCAACAAACACACCTACCGCTGC 1744
QY 1785 CAAATGGCTCTCTGTGAGCAAGGCAACCTGAGTGTGATGGGAAGACGAGCTGTAGC 1844
DB 1745 CTCAATGGCTCTCTGTGAGCAAGGCAACCTGAGTGTGACGGAAGGAGTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAACCTGTGACTGTGGCTGCGATCTCTTACCAACAGGCTGC 1904
DB 1805 GACGGCTCAGATGAGAGGACTGCGACTGTGGCTGCGGTCAITTCAGAGACAGGCTGT 1864
QY 1905 GTGTTGTGTGACAGATGCGGACGAGCGGCGAGTGGCTGCGAGTGCAGCTCCACGCC 1964
DB 1865 GTTGTGGGGCAGGATGCGGATGAGGCGAGTGGCTGCGAGTGAAGCTGCATGCT 1924
QY 1965 CTGGGCGAGGCGCATCTGTGTGGGCGCTTGCATCTCTCTGCTGCTGCTGCTGCA 2024
DB 1925 CTGGGCGAGGCGCATCTGTGGTGTCTTCCCTCATCTCTCCCACTGCTGCTGCTGCC 1984
QY 2025 GCTCATGCTTTTCAGATGACAAATTTTCAAGTACTCAGACTACAGATGTGACGCGC 2084
DB 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCACCGCAGTGCAGCGCC 2044
QY 2085 TTCTGTGGCTGTGCTGACGACGAGCAAGCGAGTGCCTCTGGGCTGCGAGGACTCAAGCTC 2144
DB 2045 TTCTGTGGCTGTGACGACGAGCGACGCGAGCGCCCTGGGGTGCAGGAGCGCAGCTC 2104
QY 2145 AACGATATCATCACCCCTTCTTCAATGATTTCACTTTCGACTATGACATCGCCTTG 2204
DB 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTTCGACTATGACATCGCGCTG 2164
QY 2205 CTGAGCTGGAGAGTGTGGAGTACAGACCGTGTGCGCCCATCTGCTGCTGCTGAT 2264
DB 2165 CTGAGCTGGAGAAACCGGAGTACAGCTTCAATGTGTGGGCGCCCATCTGCTGCGCGAC 2224
QY 2265 GCTACCATGCTTCTCTGCTGGCAAGGCCATCTGGGTTCAGAGCTGGGGGACACAAAA 2324
DB 2225 GCCTCCCATGCTTCTCTGCGGCAAGGCCATCTGGGTTCAGAGCTGGGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGAGGCTGATCTCTGAGAGGAGTGCAGTCCGTGTATCAACAGAC 2384
DB 2285 TATGGAGCACTGGCGGCTGATCTCTGCAAAAGGAGTGCAGTCCCGCTCATCAACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCCGACAGATCACCCCAAGATGATGTGTGGTTCCTC 2444
DB 2345 ACCTGTGAGGAACTCTGCGCGACAGATCACCGCGCGCATGATGTGTGGTTCCTC 2404
QY 2445 AGTGGGCTGTGAGTCTCTGCGAGGCTGACTCTGTGTGGCCCTTGTCAAGCGCGGAGAA 2504
DB 2405 AGCGGCGGTGGACTCTGTGCAAGGAGTGTTCGGGGGACCCCTGTTCAGCGTGGAGGG 2464

QY 2505 GATGGCGAATGTTCAGGCTGGTGTGCTGAGCTGGGTGAAGCTGCGCTCAGAGGAAC 2564
DB 2465 GATGGCGGATCTTCAGGCGGCTGTGTGAGCTGGGAGACGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGTACAAAGGCTCCCTGTGTGCTGGGACTGGATCAAGAGACACACTGGG 2624
DB 2525 AAGCCAGCGGTGTACAAAGGCTCCCTGTGTGCTGGGACTGGATCAAGAGACACACTGGG 2584
QY 2625 GTATAGCAGCATGTGACAGACGCGACCAAAACCCACAGGAGTGCCTGACATGCACA 2684
DB 2585 GTATAGGCGCGCGG---GCCACCCAAATGTGTACACTTGGGGGCCACCCATCTGCCACC 2641
QY 2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGTGTGGCTTCCCTCCCAACACA 2744
DB 2642 CCAGTGTGCAG-CCTGACGCTGGAGCTGGACCGCTGACTGCACACGCGCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAACCTGCATCTTAGACTCAGAGT 2780
DB 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

RESULT 11
AR229704
LOCUS AR229704 3147 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6451500.
ACCESSION AR229704
VERSION AR229704.1 GI:27269434
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3147)
AUTHORS Leon,J.W.
TITLE Imaging member containing heat switchable carboxylate polymer and method of use
JOURNAL Patent: US 6451500-A 1 17-SEP-2002;
FEATURES Location/Qualifiers
source 1. 3147
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCCCAAAACCATGGGTAGCAATCGGGCCGCGAAGCCGAGGGGCTCTCAG 104
DB 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCCGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 164
DB 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCAGAGAAAGTGAATGCTTTGAGGAA 124
QY 165 GGTGTGAGTTCCTGCTGCGCAACATGCGCAAGAAAGTGGAGAGCGAGGCCCGCGCGC 224
DB 125 GGGCTGAGTTCCTGCGCAGTCAACACGCTCAAGAAAGGTGGAAGGATGCGCCCGGGCGC 184
QY 225 TGGGTGTGCTGGTGGCAGTGTTCAGCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTG 284
DB 185 TGGGTGTGCTGGCAGCGGTGATCGGCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTG 244
QY 285 CTGCTGTGGCACTTCCATTTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCACTG 344
DB 245 CTGCTGTGGCATTTGCAGTACCGGAGCGTGTGCTCCAGAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAAATCAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
DB 305 AGGATCACAAAATCAGAAATTTTGTGGATGCTTACGAGAACTCCCAACTCCAATGAGTTGTA 364
QY 405 AGCTGGCGGACGAGTGAAGGAGGCGCTGAACTGCTGTACATGAGTCCCTGCTGCTG 464
DB 365 AGCTGGCGGACGAGGTAAGGAGCGGCTGAAGCTGTGCTGTACAGCGGAGTCCCATTTCTG 424

QY 465 GGTCCCTTACCACGAAGTGGCTGTAACTTCACTCAGTGGGCGAGTGTATCGGCTAC 524
DB 425 GGCCCTTACCACGAAGTGGCTGTAACTTCACTCAGTGGGCGAGTGTATCGGCTAC 484
QY 525 TACTGTGTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 584
DB 485 TACTGTGTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 544
QY 585 GTGAGAGAGTGTAACTTGGCAACCCCGAGAGAGAGTGTAACTTGGCTTAACT 644
DB 545 GAGAGAGAGTGTAACTTGGCAACCCCGAGAGAGAGTGTAACTTGGCTTAACT 604
QY 645 TCTGTGTGTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 704
DB 605 TCACTGTGTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 664
QY 705 AGTTTTCCTTGCATGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGT 764
DB 665 AGTTTTCCTTGCATGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGT 724
QY 765 AACAGTTCCTTACCAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGT 824
DB 725 GACAGTTCCTTACCAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGT 784
QY 825 GTGTGAGTTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 884
DB 785 GTGTGAGTTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 844
QY 885 CTGTGAGTTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 944
DB 845 CTGTGAGTTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 904
QY 945 GGCACCTTTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 1004
DB 905 GGCACCTTTCAGATTCAGATTCCTCCCAACACTGCGAGAGAGATTCAGTGGCT 964
QY 1005 ACCTGTATTAACCAATACAGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTG 1064
DB 965 ACCTGTATTAACCAATACAGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTG 1024
QY 1065 CCAAGATGAGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1124
DB 1025 CCAAGATGAGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1084
QY 1125 TACTATCCAGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1184
DB 1085 TACTATCCAGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1144
QY 1185 AACGGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 1244
DB 1145 AACGGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 1204
QY 1245 GGCTTCTGACAGAGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAG 1304
DB 1205 GGCTTCTGACAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 1264
QY 1305 CAGTTGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 1364
DB 1265 CAGTTGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 1324
QY 1365 TACAGAGAGTGGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1424
DB 1325 TACAGAGAGTGGTGTGAGAGTGGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1384
QY 1425 GGGATGTTTCAAGAGTGGTGTGAGAGAGTGGTGTGAGAGAGTGGTGTGAGAG 1484
DB 1385 GGGATGTTTCAAGAGTGGTGTGAGAGAGTGGTGTGAGAGAGTGGTGTGAGAG 1444
QY 1485 GCAGACTGCCCGAGTGTGAGAGTGGTGTGAGAGAGTGGTGTGAGAGAGTGG 1544
DB 1445 GCCGACTGCCCGAGTGTGAGAGTGGTGTGAGAGAGTGGTGTGAGAGAGTGG 1504

QY 1545 ACCTGCAAAACACAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1604
DB 1505 ACCTGCAAAACACAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1564
QY 1605 GGGGACGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 1664
DB 1565 GGGGACGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 1624
QY 1665 GGGGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1724
DB 1625 GGGGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1684
QY 1725 GAGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1784
DB 1685 GAGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1744
QY 1785 CAAAATGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1844
DB 1745 CAAAATGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1804
QY 1845 GATGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1904
DB 1805 GATGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1864
QY 1905 GTGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1964
DB 1865 GTGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1924
QY 1965 CTGGGCGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2024
DB 1925 CTGGGCGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 1984
QY 2025 GCTCATGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2084
DB 1985 GCTCATGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2044
QY 2085 TTCTGTGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2144
DB 2045 TTCTGTGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2104
QY 2145 AAAGTGTATCATGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2204
DB 2105 AAAGTGTATCATGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2164
QY 2205 CTGGGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2264
DB 2165 CTGGGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2224
QY 2265 GCTACCTGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2324
DB 2225 GCTACCTGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2284
QY 2325 GAGGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2384
DB 2285 GAGGAGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2344
QY 2385 ACCTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2444
DB 2345 ACCTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2404
QY 2445 AGTGGGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2504
DB 2405 AGTGGGGTGTGAGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTG 2464
QY 2505 GATGGGCGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2564
DB 2465 GATGGGCGAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2524
QY 2565 AAGCCAGGCGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2624
DB 2525 AAGCCAGGCGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2584
QY 2625 GTATAGCAGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAGTGGTGTGAGAG 2684

Db 2585 GTATAGGGCCGGG--GCCACCAATGTGTACACCTGCGGGGCCACCAATCTCTCCACC 2641
Qy 2585 CCTGGATACAGGAGAGAACTGACAGACATTTATGTGTGGCTCTCCCGCCCAACACA 2744
Db 2642 CCAAGTGTGACG-CCTGAGGCTGGAGACTGGACCGCTGACTGACCAAGCGCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAATGATCATCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCAGGGCTCCAAAT 2735

RESULT 12
LOCUS AR229712/c 3147 bp RNA linear PAT 20-DEC-2002
DEFINITION Sequence 18 from patent US 6451500.
ACCESSION AR229712
VERSION AR229712.1 GI:27269442
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3147)
AUTHORS Leon, J.W.
TITLE Imaging member containing heat switchable carboxylate polymer and method of use
JOURNAL Patent: US 6451500-A 18 17-SEP-2002;
FEATURES Location/Qualifiers
source 1..3147
BASE COUNT 583 a 958 c 952 g 654 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAG 104
Db 3143 GAGCGGCTCGGGGTACCATGGGAGCGATCGGCCCGCAAGGCGGAGGGGCCGAG 3084
Qy 105 GACTTCGGCGGGGACTCAAGTCAACTCCCGGTAGAGAACATGAATGGCTTTGAGAG 164
Db 3083 GACTTCGGCGGGGACTCAAGTCAACTCCCGGCAGAGAAAGTGAATGGCTTGGAGAA 3024
Qy 165 GGTGTGAGTTCTCGCTCGGCAATGCCAGAAAGTGGAGAGCGAGCCCCCAGCGC 224
Db 3023 GGCCTGGAGTTCTCGCCAGTCAACAGCTCAAGAAAGGTGGAAGAGCATGCGCGGGCGC 2964
Qy 225 TGGGTGTGTGTGGGAGTGTCTTTCAGCTTCTCTTGTCTCTCTCTCTCTCTCTCTCTCT 284
Db 2963 TGGGTGTGTGTGGGAGTGTCTTTCAGCTTCTCTTGTCTCTCTTGTCTTGTCTTGTCT 2904
Qy 285 CTGTGTGGCACTTCATATGCGAATATGCGGATGCGGGTTCAAAAGTCTTCAATGGCCATCTG 344
Db 2903 CTGTGTGGCACTTCATATGCGAATATGCGGATGCGGGTTCCAGAGGTCTTCAATGGCTACATG 2844
Qy 345 AGGATCAAAATGAGATCTTCTGTGATGCGTATCAGAACTCCACTCCACAGAGTTTATC 404
Db 2843 AGGATCAAAATGAGATCTTCTGTGATGCGTATCAGAACTCCACTCCACTGAGTTTGTGA 2784
Qy 405 AGCCTGGCCAGCCAGGTGAAGAGGCGCTGAAGTGTCTGTACAAATGAATGCCCTGTCTCTG 464
Db 2783 AGCCTGGCCAGCAAGGTGAAGAGGCGCTGAAGTGTCTGTACAGCGGAGTCCCAATCTCTG 2724
Qy 465 GGTCTCTACCAAGAGAGTGGCTGTAACTGCTTTCAGTGGGAGGAGTGTATCGGCTAC 524
Db 2723 GGCCCTTACCAAGAGAGTGGCTGTAACTGCTTTCAGTGGGAGGAGTGTATCGGCTAC 2664
Qy 525 TACTGGTCAGAGTTTCAGATCCCGCCACCTGCGCAGAGAGGTGTATCGGCTAC 584
Db 2663 TACTGGTCAGATCCCGCCACCTGCGCAGAGAGGTGTATCGGCTAC 2604

Qy 585 GTGAGCGAGTTGTAACTTTGCCACCCCGGACGCGGACATGAAATCCTTCGTCTCAACA 644
Db 2603 GAGAGCGGTAGTGTATGCTGCCCCCGGGCGGCTCCCTGAAAGTCTTTGTGGTCAAC 2544
Qy 645 TCTGTGTGGTCCCTTTCCCACTTTGACCCCGAGAAATGTGACAGAGACTCAGGACCAACAGCTGC 704
Db 2543 TCAGTGTGTGGCTTTCCCGCCAGGACTCCAAACAGTACAGAGGACCCAGGACCAACAGCTGC 2484
Qy 705 AGTTTGTGCTGATGCGCATGTGTGACAGAGTGCACGCTTCTACTACCTCGGTGGCTTCCCC 764
Db 2483 AGCTTTGGCCTGCACGCCCGCGGTGTGAGCTGTATGCGCTTACACAGCCCGGGTTCCT 2424
Qy 765 AACAGTCCCTACCGCGCGCATGCCCGCTGACAGTGGGTCTCTGCGGGGAGCGCGACTCT 824
Db 2423 GACAGCCCTTACCCCGCTCATGCCCCGTGCGAGTGGGCCCTTGGGGGGAGCCCGACTCA 2364
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTCCCTGTGTGATGAGCATGGCAGTGAC 884
Db 2363 GTGCTGAGCCTCACCTTCCGAGCTTTGACCTTGTGCTGCTGCGACGAGCGCGCAGCGAC 2304
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCATGGAACCCCAAGCTGTGTGTGCGGTGTGT 944
Db 2303 CTGGTCACCGTGTATGATAGCCTGAGCCCATGGAACCCCAAGCTGTGTGTGCGGTGTGT 2244
Qy 945 GGCACCTTCTCACCTCTCTACAACTGACTTTCTCTCTCTCCAGAAAGCTCTTCTCTGTG 1004
Db 2243 GGCACCTTCT 2184
Qy 1005 ACGCTGATAACCAATATCTGACCGCGGACATCTCTGGCTTTGAGGCCACTTTTCTTCCAGCTG 1064
Db 2183 ACATGATAACCAACACTGAGCGCGGACATCTCTGGCTTTGAGGCCACTTTTCTTCCAGCTG 2124
Qy 1065 CCCAAGATGAGAGCTGTGGGGCTTTTGTGAGTGACACCCCAAGGGACATTTAGAGCGCCC 1124
Db 2123 CTAAGATGAGAGCTGTGGGGCGCTTACGTAAAGCCCAAGGGACATTTCAACAGCGCCC 2064
Qy 1125 TACTATCCAGCGCCACTACCGCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 2063 TACTATCCAGCGCCACTACCGCGCCCAACATTTGACTGACATGGAATATTTAGGTGCCCAAC 2004
Qy 1185 AACCGAACTGGAAGTGTGGCTTCAAACTTTCTATCTGTGTGACCCCAACGATCAAGTGT 1244
Db 2003 AACCGAACTGGAAGTGTGGCTTCAAACTTTCTATCTGTGTGAGCGCGGTGCTGCG 1944
Qy 1245 GGTCTCTGACCAAGGACTATGTGGAGTCAACGGGGAGAGTACTGGGTGTGAGAGTTC 1304
Db 1943 GGCACCTTCCCGCCCAAGGACTACGTGGAGTCAATGGGGAGAAATCTCGGAGAGAGTTC 1884
Qy 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAACAGTCCACTTCCACTTCTGATCACTCG 1364
Db 1883 CAGTTTGTGTGAGCAGCAACAGCAGCAACAGATCAAGTTCGCTTCCACTCAGATCACTCC 1824
Qy 1365 TACAGGACACCGGTTCTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
Db 1823 TACAGGACACCGGTTCTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1764
Qy 1425 GCGATGTTCATGTGCAAGTGTGACGCTGCATCCGAAGGAACTGCGCTGCGAGCGGTGG 1484
Db 1763 GCGAGTTTACGTGCGCGCAGCGGGCGGTGTATCCGGAAGAGCTGCGCTGTGATGGCTGG 1704
Qy 1485 GCAGCTGCGCGGATTTATGATGAGCGGTACTGCGGATGCAATGCCACCCACAGTTC 1544
Db 1703 GCGGCTGCGCGGACCGACAGCGATGAGTCAACTGCAAGTTGCGACGCGCGCCACAGTTC 1644
Qy 1545 ACGTGCAAAACCAAGTTCTGCAAGCGCTCTTCTGTGGTCTGTGACAGTGTCAAGACTGT 1604
Db 1643 ACGTGCAAAACCAAGTTCTGCAAGCGCTCTTCTGTGGTCTGTGACAGTGTCAAGACTGT 1584
Qy 1605 GCGGAGGAGTGCAGGAGGGCTGCGCTGCTCTGCTGGGAGTTTCAAGTGTTCAT 1664
Db 1583 GCGGAGGAGTGCAGGAGGGCTGCGCTGCTCTGCTGGGAGTTTCAAGTGTTCAT 1524
Qy 1665 GCGGAGTGTCTCTCTCTCAGAGCCAGAGTGTATGGGAAGGACAACTGTGGAGATGGGTCT 1724

Db 1523 GGAAGTGCCTCTCGAAAGCCAGCAGTGAATGGGAAGACGACTGTGGGGACGGGTCC 1464
Qy 1725 GAGAGGCTTCATGTGACAGCGTGAATGTCTCTTGGACCAATATACCTACCGCTGC 1784
Db 1463 GAGAGGCTTCCTGCCCAAGGTGAACGTGTCTCACTTGTACCAACACACCTACCGCTGC 1404
Qy 1785 CAAAATGGCCCTCTGTCTGAGCAAGGGCAACCCCTGAGTGTGATGGGAAGACGAGCTGTAGC 1844
Db 1403 CTCAATGGGCTCTGCTTGTAGCAAGGGCAACCCCTGAGTGTGACGGGAAGGAGCTGTAGC 1344
Qy 1845 GATGGCTCCATGAGAAAACTGTGACTGTGGGTGGATGCTCTTTTACCAACAGGCTGCG 1904
Db 1343 GACGGCTCAGATGAGAAAGGACTGCACTGTGGGTGGGTGCGGTCTTACGAGACAGGCTCGT 1284
Qy 1905 GTGGTGGTGGCCAGAAATCGGACAGGGGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db 1283 GTTGTGGGGCCAGGATCGGATGAGGGGAGTGGCCCTGGCAGGTGAGCTGATGCT 1224
Qy 1965 CTGGGCCAGGCCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1223 CTGGGCCAGGCCCACTCTCGGTGCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGOC 1164
Qy 2025 GCTCATTTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGAAGGCC 2084
Db 1163 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAAGCTGGAAGGCC 1104
Qy 2085 TTCTGCGGTCTGTGGACCAAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 1103 TTCTGCGGTTCGACGACAGCGAGCGAGCCCTCTGGGTGCAGAGCGCAGCTC 1044
Qy 2145 AAAGTATCATACCCACCTCTTCAATGATTTACCTTGCATGATGATGATGATGATGATG 2204
Db 1043 AAGCGCATCATCTCCCAACCCCTCTTCAATGACTTCACTTTCGACTATGATGATGATG 984
Qy 2205 CTGAGCTGAGAGAGTGGGTGAGTACAGCAGCGTGTGGCGGCCCACTGCTGCTGCTGAT 2264
Db 983 CTGAGCTGAGAGAACCGGACAGAGTACAGTCCATGTTGGTGGCGGCCCACTGCTGCTGCGGAC 924
Qy 2265 GCTACCATGCTCTTCCCTGTGGCAAGGCCATCTGGGTGCAGAGCTGGGGGCGCACAAAA 2324
Db 923 GCCTCCCATGCTTCTCCCTGCGGCAAGGCCATCTGGGTCAAGGGTGGGGACACCCAG 864
Qy 2325 GAGGAGGTACCGAGCGCTGATCCTCGAGAGGTGAGATTCGTGTATCAACACGAGCC 2384
Db 863 TATGAGGCGCATGCGCGCTGATCCTGCAAAAGGCTGAGATCCGCGTCAATCAACAGACC 804
Qy 2385 ACCTGTGAGGACCTCATGCGCGCAGCAGATCACCCACGAATGATGTGTGGGTTCCTC 2444
Db 803 ACCTGCGAGAACCTTCTGCGCGCAGCAGATCACGCCGCGCATGATGCGTGGGCTTCCCTC 744
Qy 2445 AGTGGGGGTGTGACTCTCTGCCAGGGTGACTCTGTGTGGCCCTTGTCAAGCGGGAGAAA 2504
Db 743 AGCGGCGCGTGGACTCTGCGAGGGTGAATTCGGGGGACCCCTGTCCAGCGTGGAGCG 684
Qy 2505 GATGGGCGAATGTTCCAGGCTGTGTGGTGTGAGCTGGGGTGAAGCTCGCTCAGAGAAC 2564
Db 683 GATGGGCGAATCTTCCAGGCGCGTGTGTGTGAGCTGGGGGAGACGCTCGCTCAGAGAAC 624
Qy 2565 AAGCCAGCGGTGTACAAAGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 623 AAGCCAGCGGTGTACAAAGCTCCCTGTGTTCGGGACTGGATCAAGAGAACACTGGG 564
Qy 2625 GTATAGCAGCATGAGACAGACGCGACCAAAACACCCACAGGGATCCCGACATGACCA 2684
Db 563 GTATAGGGGCGCGG- -GCCACCCAAATGTGTACACCTGGGGGCCACCCCATGCTCCACC 507
Qy 2685 CCTGATACAGGAGGAGAACACTGACGACATTTATGTGTGGCTTCCCGCTCCCGCCCAACACA 2744
Db 506 CCAGTGTGCAGC-CTTGAGGCTGGAGACTGGACCGCTGACTGACACAGCGGCC-CCAGA 449
Qy 2745 ACCCAGACTGTGAACCTGATCCTTATGAGACTCAGAGT 2780

448 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAT 413
RESULT 13
AX207897
LOCUS 3147 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 1 from Patent WO0157194.
ACCESSION AX207897
VERSION AX207897.1 GI:15422495
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the encoded proteins and methods based thereon
JOURNAL Patent: WO 0157194-A 1 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
1. 3147
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="Nucleotide encoding MTSPL"
23. .>2589
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC60377.1"
/db_xref="GI:15422496"
/translation="MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNV
KVEKHGPRVWVLAALVLLIGLFLVHLYQDRVVRQKVFNGVYRITNENFV
DAYENSGSTFVSLASKVDALLKLYSGVFLGPKHESAVTAFSGSVIAYWSFS
IQHIVLEAERVMABERVVMLPPARSLAKSVVTSVAFTDSTKVQRTQDNCSFGL
HARGVLEFTTTPGPDSPYBAHARCOWALRGDADSVLSLTPRSPDLASDERGSDIV
TYNLSLMEPHALVOLCGTTPSYNLTTHSSONVLLITLITNTERHGFERTFQL
PRMSSCGGRKRAQGTFFSPYYPGHYPNIDCTWNIENVPNNQHVKSFKFFLLYPSG
PACTCPKDYVEINGEKYCGERSQFVVTSSNKITVRFHSDQSYDTGTFLAELYSYS
DPCPGQCTRTGCRKELRCGWADCTDHSDELNCSDCAGHOFCKNFKCPKPVVVT
DSVNDGDNSSDGGSCPAQTCRNSGKCLSKSQCCGKDCGDSDEASCPRKNNVT
CTKHYRCLNGLCISKNPECDGKDCSDGSEKDCDCGLRSTFRQARVVVGTDABEG
EMQVSVHALGGHICGASLISPNWLVAACHTIDDRFRYSIDPTQWTAFLGLHQDS
QAPQVQBRRLRIIISHFFNFDFDYDIALLELEKPAEYSVMVPICLPDASHVFP
AKGAIWVTGWHTQGTGALILQKREIVNTCCENTLIPQOITPRMCMVGFSLSGV
DSQCGSGGLPSVZADGRIFQAGVVSMDGCAQRNKPQVYTRLPFRDWIKENTGV"
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN
Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
Qy 45 GATCGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCCGAGGGGCTCTCAG 104
Db 5 GAGCGCCTCTGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCCGAAG 64
Qy 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCACAGAGAGTGAATGGCTTTGAGGAA 124
Qy 165 GGTGTGGAGTTCCTGCTCGGAACAATGCCAGAAAGTGGAGAGCGAGGCCCGCCAGCGC 234
Db 125 GCGTGGAGTTCCTGCGCAGTCAACACAGTCAAGAGGTGGAAGAGCATGGCCCGGCGC 184
Qy 225 TGGGT 284
Db 185 TGGGT 244
Qy 285 CTGGTGTGGCATTTCATTATCGAATGTGGGGTTCAAAAAAGCTTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCATTTCAGTACCGGAGCGTGGGTGTCAGAGAGCTCTTCAATGGCTACATG 304

Db 2465 GATGGGGGATCTTCACAGGCGCGTGTGGTGAAGTGGGAGACGGCTGCGTCAAGAGAAC 2524
QY 2565 AAGCCAGGCGGTGTACACAAAGCTCCCTGTAGTTGGGACTGGATCAAGAGACACTGGG 2624
Db 2525 AAGCCAGGCGGTGTACACAAAGCTCCCTGTAGTTGGGACTGGATCAAGAGACACTGGG 2584
QY 2625 GTATAGCAGATGACAGACAGCGGACCAAAACACCCAGGATGCCGACATGACACA 2684
Db 2585 GTATAGGCGGCGGG--GCCACCCAAATGTGTACACCTGGGGGCCACCCATGTCACC 2641
QY 2685 CTTGGATACAGGAGAGAACACATGACACATTTATGCTGTGGCTCCCGCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGACGCTGGAGCTGAGCCGCTGACTGACACGCGCC-CCAGA 2699
QY 2745 ACCAGACTGTGAAGTGCATCTCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 14
AX207945
LOCUS 3147 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 49 from Patent WO0157194.
ACCESSION AX207945
VERSION AX207945.1 GI:15422543
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the encoded proteins and methods based thereon
JOURNAL Patent: WO 0157194-A 49 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source 1..3147
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
CDS <1865..2590
/notes="protease domain of MTSpl"
/codon_start=1
/protein_id="CAC60383.1"
/db_xref="GI:15422544"
/translation="VVGSTDADEGWVQVSLHALGQGHICGASLISPNMLVSAHCY
IDRGRFYSPTQWTFPLGLHDOSRPAQVQRRLKRIISHPFNDFDYDIALLE
LEKPAEYSMVRPILCPDASHVFPAGKAIWVTGWHTQYGTGALILQKEIRIVNT
TCENLLPQOITPRMCMVGLSGVDSGSGPLSSVEADGRIFQAGVVSWDGCAQ
RNRPGVTRLPFLRWIKENTGV"
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.28; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATGGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGCGGAGGGGCCGGAAG 64
QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCGGGTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCGGGCACGAGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGAGTTCTCGCTCGGACAAAGTCCCAAGAACTGGAGAGCGAGCCCGGCGCGC 224
Db 125 GGCCTGAGTTCTTCGCCAGTCAACAGTCAAGTCCCGGCACGAGAAAGTGAATGGCTTTGAGGAA 184
QY 225 TGGTGTGCTGTGGTGGGAGTGTCTTTCAGCTTCTCTTGTCTCCCTCATGCGGTGGCTTG 284
Db 185 TGGTGTGCTGTGGGAGCGGTGCTGATCGGCTCTCTTGTCTTGTGGGATCGGCTTC 244

QY 285 CTGGTGTGGCACTTCCATTATCGAAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCACTTTCAGTACCGGGACGTGCGTGTCCAGAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCAAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCAAAAATGAGAAATTTTGTGATGCTTACAGAACTCCAACTCCACTGAGTTTGTGA 364
QY 405 AGCTCGGCGAGCCAGTGAAGGAGCGCTGAACTGCTGTAGAACTGAAGTCCCTGCTCTG 464
Db 365 AGCTCGGCGAGCAAGTGAAGGAGCGCTGAACTGCTGTAGAGCGAGTCCCATTCCTG 424
QY 465 GGTCCCTTACCACAAAGAGTGGCTGTAACTTCCCTTCACTGAGGCGAGTGTCACTCGCTAC 524
Db 425 GGGCCCTTACCACAAAGAGTGGCTGTGAGCGCTTCCAGGAGGCGAGCTCATCGCTAC 484
QY 525 TACTGTGAGAGTTTCAAGATCCCCCACACCTGGGCAAGAGAGTTGATCGCGCATGGCT 584
Db 485 TACTGTGCTGAGTTTCAAGATCCCCCACACCTGGGAGGAGGCGGCGCTCATGGCC 544
QY 585 GTGGAGCGAGTTTGAACATTTGCCACCCGAGCACGGGCACTGAAATCCCTTCTGTCTAAACA 644
Db 545 GAGGAGCGGCTGATGCTGCTGCGGCGCGCTCCCTGAACTCTTTTGGTCAACC 604
QY 645 TCTGTGTGGCTTCCCTTCCCACTTGAACCCAGAAATGCTGCAGAGGACTCAGGACAAAGCTGC 704
Db 605 TCAGTGTGGCTTTCCTCCCACTGCAAAAACAGTACAGAGGACCCAGGACAAAGCTGC 664
QY 705 AGTTTGGCTTGCATGCGCATGGTGCAGAGTGCAGAGTTCATACCTCCCTGCTTCCCT 764
Db 665 AGTTTGGCTTGCATGCGCATGGTGCAGAGTGCAGAGTTCATACCTCCCTGCTTCCCT 724
QY 765 AACAGTCCCTTACCCGCGCATGCGCTGCGAGTGGTCTGCGGGGGAGCGCGACTCT 824
Db 725 GACAGCGCTTACCCGCTCATGCGCGCTGCGAGTGGGCGCTGCGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCTTCACTTCCGAAAGCTTGTGATGCTGCTCCCTGTGATGAGCAATGAGTGC 884
Db 785 GTGCTGAGCTTCACTTCCGAGCTTGTGAGTGTGCTGCTGCTGCGAGCGCGGCGAGCGAC 844
QY 885 CTGGTCACTGCTGATGAGCTGAGCCCTGAGCCCATGAAACCCAGCTGCTGCTGCGGCTGTGT 944
Db 845 CTGGTCACTGCTGATGAGCTGAGCCCTGAGCCCATGAGCCCATGAGCCCTTGTGCTGCTGTGT 904
QY 945 GGCACCTTCTTCACTTCCCTTACAACTGACTTCTCTCTCCCGAGAACCTTCTTCTTGTGTC 1004
Db 905 GGCACCTTCTTCACTTCCCTTACAACTGACTTCTCTCTCCCGAGAACCTTCTTCTTGTGTC 964
QY 1005 AGCTGATTAACATGACTGACCGGCGACATCTGCTGCTTTGAGGCGCACTTCTTCTTCAAGCTG 1064
Db 965 ACATGATTAACCAACACTGAGCGCGCGCATCCCGGCTTTGAGGCGCACTTCTTCTTCAAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGTGCGGCTTTTGTGAGTGACACCCCAAGGAGACATTTAGCAGCCCC 1124
Db 1025 CCTAGATGAGCAGCTGTGAGGCGCTTACGTAAGACCCAGGAGCAATTCACAGCCCC 1084
QY 1125 TACTATCCAGGCGCACTACCCCGCCAACTCACTGACATGCAATGCAATATCAAGTGGCCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCCCGCCAACTGACTGCAATGCAATGCAATGAGTGGCCCAAC 1144
QY 1185 AACCGAAAGTGAAGTGGCTTCAAACTTCTTATCTGCTGAGGAGCGGCAAGCTTACAGTGC 1244
Db 1145 AACCGAAAGTGAAGTGGCTTCAAACTTCTTATCTGCTGAGGAGCGGCGCTTCTTCTTCAAGCTG 1204
QY 1245 GGTCTCTGACCAAGGACTATGTGGAGATCAACGGGAGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCGGCGAGGACTGCTGGAGATCAATGGGAGAAATFACCTCGGAGAGAGGTCC 1264
QY 1305 CAGTTTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTCTGATCACTG 1364
Db 1265 CAGTTTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCACTTCCACTCAGATCAGTCC 1324

QY	1365	TACACGGACACCGGTTCTTAGCTAGTACTCTCTCTACGACTCCAAACGCCGCTGCCCA	1422
Db	1325	TACACGCACACCGGCTCTTTAGCTGTAATACCTCTCTACGACTCCAGTACCCATGCCCG	1384
QY	1425	GGGATGTTCTATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGGCGTGCAGCGCTGG	1484
Db	1385	GGGCAGTTACGTCGCCGACCGGGCGGTGTATCCGGNAGAGCTGGCTGTGATGGCTGG	1444
QY	1485	GCAGATGCCCGGATTAATGATGATGAGCGTTACTGCGGATGGAATGCACCCACCAAGTTC	1544
Db	1445	GCCGACTGCACCGACACACGCGATGAGCTCAACTGCAGTTGCGACGCCGCGCCACCAAGTTC	1504
QY	1545	ACGTGCAAAAACAGTTCTCAAGCCCTCTTCTGGGTCGTGCACAGTGTCAACGACTGT	1604
Db	1505	ACGTGCAGAAACAGTTCTCGAAGCCCTCTTCTGGGTCGTGCAGACTGTGAAAGACTGC	1564
QY	1605	GGGACGGAAAGTGAACGAGGAGGCTCAGCTGTCCTCTGGGATTTCAAGTGTCCAAAT	1664
Db	1565	GGAGACAAACGACACGACGAGGGGTGCAGTTGTCCGGCCGAGACCTTCAGGTGTCCAAAT	1624
QY	1665	GGGAAGTGTCTCCCTCAGAGCCGAGAGTGTAAATGGGNAGACAACTGTGGAGATGGGTCT	1724
Db	1625	GGGAAGTGTCTCTCGAAAGCCAGCAGTGCAAATGGGAAGACCACTGTGGGACGGGTCC	1684
QY	1725	GACGAGCTTCATGTGACAGCGTGAATGTGCTCTCTTGCACCAAAATATACCTACCCTGC	1784
Db	1685	GACGAGCTCTCTGCCCAAAGGTGACGTGCTCACTTGTACCAACACACCTACCGCTGC	1744
QY	1785	CAAAATGGCTCTGTCTGACCAAGGCGAACCTGAGTGTGATGGGAGACGACTGTAGC	1844
Db	1745	CTCAATGGGCTCTGCTTGTAGCAAGGGCAACCTCTGAGTGTGACGGGAGGAGGACTGTAGC	1804
QY	1845	GATGGGTCCTCAGAGAAAACTGTGACTGTGGGCTCGGATCTCTTTACAAACAGGCTCGC	1904
Db	1805	GACGGCTCAGATGGAAGGACTCGGACTGTGGGCTGGGTCATTACGAGACAGGCTCGT	1864
QY	1905	GTGGTTGGTGACGAATGCGGACGAGGGCGAGTGGCCCTTGGCAGGTGAGCTTCCACGCC	1964
Db	1865	GTGTGTTGGGGGACCGATGCGGATGAGGGCGAGTGGCCCTTGGCAGGTAAAGCTTGCATGT	1924
QY	1965	CTGGGCCAGGGCCACTTGTGTGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA	2024
Db	1925	CTGGGCCAGGGCCACATCTCGGTGCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGCC	1984
QY	2025	GCTCATGTCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGGACGGCC	2084
Db	1985	GCACACTGCTACATCGATGACAGAGGANTCAGGTACTCAGACCCACGCGAGTGGACGGCC	2044
QY	2085	TTCTCTGGGCTGTGGACACAGACGCGAGTGCCTCTGGGCTGCAGGAGCTGAAGCTC	2144
Db	2045	TTCTCTGGGCTGTGCACACACAGACCGCAGCGCCCTTGGGCTGCAGGAGCGCAGGCTC	2104
QY	2145	AAACGTATCATCACCCACCTTCCTTCAATGATTTACCTTTCGACTATGACATCGCCTTG	2204
Db	2105	AAGCGCATCATCTCCACCCCTTCTTCAATGACTTTCACCTTTCGACTATGACATCGCGCTG	2164
QY	2205	CTGAGCTGAGAACTCGGTGAGTACAGACCGTGTGTGGCCCCCATCTGTGCTGCTGTGAT	2264
Db	2165	CTGAGCTGGAGAAACCGGCAGAGTACAGTCCATGTTGGGCCCATCTGTCTGCGCGAC	2224
QY	2265	GCTTACCCATGCTCTTCCCTGTGCAAGGCCCATCTGGGTACAGGCTGGGGGCACACAAA	2324
Db	2225	GCCTCCCATGTCTTCCCTGTCCGGCAGGCCCATCTGGGTCACTGGGCTGGGACACACCCAG	2284
QY	2325	GAGGAGGATACCGGAGCGCTGATCTCTGCAAGAGGTTAGATCCGTGTATCAACACGACC	2384
Db	2285	TATGGAGGCACTGGCGCGCTGATCTTGCAAAAGGGTGAGATCCGCGCTCATCAACACGACC	2344
QY	2385	ACCTGTGAGACCTCATGCCGACGAGATCACCCCAAGATGATGTGTGGGTGTTTCCTC	2444
Db	2345	ACCTGCGAGAACCTCTTGGCCGACGAGATCACGCCGCGCATGATGTGCGTGGGCTTCTCTC	2404
QY	2445	AGTGGGGGTGTGAACCTCCTGCCAGGGTGACTCTGTGGCCCTTGTCAAGCGCGGAGAAA	2504

Db	2405	AGCGGGCGGTGACTCTCTGCCAGGCTGATTCCTGGGGGACCCCTGTCCAGCGTGGAGCG	2464
QY	2505	GATGGCGGAATGTTCCAGAGCTGTGTGTGCTGAGCTGGGTGTAAGCGCTGCGCTCAGAGGAAC	2564
Db	2465	GATGGCGGATCTTCCAGCCCGTGTGGTGTGCTGGGAGACGGCTGCGCTCAGAGGAAC	2524
QY	2565	AAGCCAGGCGGTGTACACAAGGCTCCCTGTAGTTTCGGGACTGGATCAAGAGCACACTGGG	2624
Db	2525	AAGCCAGGCGGTGTACACAAGGCTCCCTGTGTTCGGGACTGGATCAAGAGGAACACTGGG	2584
QY	2625	GTATAGCAGCATGACAGACAGACCGGACCAACACCCACCCACAGGATGCCGACATGCACA	2684
Db	2585	GTATAGGCGCGGG---GCCACCCAAATGTGTACACTCTGGGGGCGCACCATGTCTCACCC	2641
QY	2685	CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCTCTCCCCCCCCCAACACA	2744
Db	2642	CCAGTGTGCAGC-CTGTGAGGCTGGAGACTGGACCGCTGACTGCCACGAGCGCC-CCAGA	2699
QY	2745	ACCCAGACTGTGAATGCATCCTTAGGACTCAGAT	2780
Db	2700	ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT	2735
RESULT 15			
AF057145			
LOCUS	AF057145	3147 bp mRNA linear PRI 22-SEP-2000	
DEFINITION	Homo sapiens serine protease TADG15 mRNA, complete cds.		
ACCESSION	AF057145		
VERSION	AF057145.1	GI:10257389	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Tanimoto, H., Underwood, L.J., Wang, Y., Shigemasa, K., Parmley, T.H. and O'Brien, T. J.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-APR-1998) Biochemistry, University of Arkansas for Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205, USA		
FEATURES	Location/Qualifiers		
source	1..3147		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
CDS	23..2590		
	/notes="associated with human ovarian cancer"		
	/codon_start=1		
	/product="serine protease TADG15"		
	/protein_id="AAG15395.1"		
	/db_xref="GI:10257390"		
	/translation="MGSDRARKGLGGGPKDFAGLKYNGRHKVNGLEGVFLPYNVNV KVEKHGPRWVLAALVLLGLVGLFVHLQYRDVRYQKVFNGVMRTNENFV DAYENSNTEFVSLASKVKDALKLYSGVPFLGPHKESAVTAFSEGSVIAYWEFS I-QHVELMRVWASERVVMPAPRSLKSFVTVSWAFFPTDSKTVQRTQDNCSGDL HARGVELMRFTPGPSPYPAHRCQWALRGDADVLSLTPRSDLASCDRSGDGL TYNTLSMPERHALVQLCTPYPSYNLTFHSSQNVLLTLITNTPRRHPGFATFPQL PMSSCGGRRLKRAQCTFNSYPYPHPNDCTWNIENPNQHKVSKFFVLLPVGV PACTPKDVELNGEKYCCERSQFVTVNSNKITVRFHSDQSYTDTGLFALYSYDSS DPQCQCTCTGCTCKELRCQGMADCTDSDDELNCSQDAGHQFTCKNFKCFLPFWIC DSVNDDGNSDFQSGSCPAQTPRCSNGKLSKQCNCKDDCGSDSDENASCFAVNVIT CKTYRCLNGLCKNPECDGKDCDGLSRFTQKRVVGGTDDADEG EWPVQSHALUGQGHICGASLISPNWLVSAAHYIDDRGFRYSFTQKRVVGGTDDADEG ORSAFVTEGRLRKLIIISHFPFNDFDYDIALLELEKPAEYFSVMVRPCLPDSAHVFP AGKATVGTGWGTQYGGTGLLQKEIRVINTQTCENLLPQITPRMCMVGLSGGV DSCQSDSGGSLSSVADGRIFOAGVVSMDGCAQRNKPVGVTFLPFRDWIKENTGV"		
BASE COUNT	654 a	952 c	958 g 583 t
ORIGIN			
Query Match	60.6%;	Score 1883.2;	DB 9; Length 3147;
Best Local Similarity	81.2%;	Pred. No. 0;	
Matches 2231:	Conservative	0: Mismatch	508: Indels
			5: Gaps

QY 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGCAAGCGCGGAGGGGCTCTCAG 104
DB |||||
5 GAGGGGCTCGGGGTACCATGGGGAGGATCGGGCGCAAGGGCGGAGGGGCGCGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTAACTCCCGCTAGAGAACTAATGCTTTGAGGAG 164
DB |||||
65 GACTTCGGCGCGGACTCAAGTAACTCCCGCTAGAGAACTAATGCTTTGAGGAA 124
QY 165 GGTGTGGAGTTCCTGCTCGGACCAATGCAAGAAAGTGGAGAGCGAGCGCCAGGCGC 224
DB |||||
125 GGGTGGAGTTCCTGCTCGGACCAATGCAAGAAAGTGGAGAGCGAGCGCCAGGCGC 184
QY 225 TGGGTGGTGTGGTGGGAGTCTGTTGAGTCTCTCTGCTCTCCCTCATGCTGGCTTG 284
DB |||||
185 TGGGTGGTGTGGGAGCGGTGATCGGCTCTCTGCTCTGCTGGGATCGGCTTC 244
QY 285 CTGGTGGGACTTCATTAATCGGAATGGGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
DB |||||
245 CTGGTGGGACTTCATTAATCGGAGCGGTGTCAGAAAGTCTTCAATGGCTACATG 304
QY 345 AGGATCAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTATC 404
DB |||||
305 AGGATCAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTATC 364
QY 405 AGCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGACAAATGAAGTCCCTGCTCTG 464
DB |||||
365 AGCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGACAGCGAGTCCCATTCCTG 424
QY 465 GGTCTCTACCAAGAGTGGCTGTAAGTCTGCTGACAAATGAAGTCCCTGCTCTG 524
DB |||||
425 GGGCCCTTACCACAGGAGTGGCTGTGACGGCCCTTACAGAGGCGAGCGTCACTGCGCTAC 484
QY 525 TACTGTGACAGTTCAGATCTCCCGCCACACCTGCGAGAGAGTGTATCGCGCCATGGCT 584
DB |||||
485 TACTGTGACAGTTCAGATCTCCCGCCACACCTGCGAGAGAGTGTATCGCGCCATGGCT 544
QY 585 GTGGAGGAGTGTAACTTGCACCCCGAGCAAGGCACTGAAATCTCTGCTGCTAAACA 644
DB |||||
545 GAGGAGCGGCTAGTCACTGCTGCGCCCGCGGCGGCTCCCTGAACTCTTTGTGGTCAAC 604
QY 645 TCTGTGGTGGCTTCCCATGAGCCCAAGTATGCTGAGAGGACTCAGAGCAACAGCTGC 704
DB |||||
605 TCAGTGGTGGCTTTCGCCACGACTCCAAACAGTACAGAGGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTGGCTGCAATGCTGAGCAGTGCACAGCTTCACTACCTGCTGCTTCCCTCC 764
DB |||||
665 AGCTTTGGCTGCAATGCTGAGCAGTGCACAGCTTCACTACCTGCTGCTTCCCTCC 724
QY 765 AACAGTCCCTPACCGGCGCATGCCGCTGCGAGTGGGTCTGCGGGGAGCGCCGACTCT 824
DB |||||
725 GACAGCCCTTACCCCGCTCATGCGCGCTGCGAGTGGGCGCTGCGGGGAGCGCCGACTCA 784
QY 825 GTGCTGAGCCTTACCTTCGAGCTTGAATGCTGCTCCCTGATGAGCATGCGAGTGAC 884
DB |||||
785 GTGCTGAGCCTTACCTTCGAGCTTGAATGCTGCTCCCTGATGAGCATGCGAGTGAC 844
QY 885 CTGGTCAACCGTGTATGATGCTGAGCCCATGAAACCCCAACGCTGTGCTGGGCTGTGT 944
DB |||||
845 CTGGTCAACCGTGTATGATGCTGAGCCCATGAAACCCCAACGCTGTGCTGGGCTGTGT 904
QY 945 GGCACTTCTACCCCTCTCAACCTGAGTCTTCTCTCTCCAGAAAGTCTTCTGCTGTC 1004
DB |||||
905 GGCACTTCTACCCCTCTCAACCTGAGTCTTCTCTCTCCAGAAAGTCTTCTGCTGTC 964
QY 1005 ACCGTGATACCAATCTGAGCGGCGCATCTGCTGCTGCTGAGGCGCTTCTTCTGAGCTG 1064
DB |||||
965 ACCTGATACCAATCTGAGCGGCGCATCTGCTGCTGCTGAGGCGCTTCTTCTGAGCTG 1024
QY 1065 CCCAAGTATGAGCTGTGGCGGCTTTTTCAGTGACACCCCAAGGAGCATTTAGCAGCCCC 1124
DB |||||
1025 CCTAGGATGAGCTGTGGAGCGCGCTTACGTAAAGCCCAAGGGGACATTTCAACAGCCCC 1084

QY 1125 TACTATCCAGGCGACTACCGCGCCAAACATCACTGCACATGGAATATCAAGGTGCCCAAC 1184
DB |||||
1085 TACTATCCAGGCGACTACCGCGCCAAACATCACTGCACATGGAATATCAAGGTGCCCAAC 1144
QY 1185 AACCGGAAAGTGAAGTGGCTTCAAACTCTTCTATCTGGTGACCCCAACGTAACAGTG 1244
DB |||||
1145 AACCGGAAAGTGAAGTGGCTTCAAACTCTTCTATCTGGTGACCCCGCGTGGCTGCG 1204
QY 1245 GGCTCTCTGACCAAGGACTATGTGGAGATCAAGGGGAGAGTACTGCGGTGAGAGGTCC 1304
DB |||||
1205 GGCACTGGCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGATTAAGTCCACTTCCATTTCTGATCACTCG 1364
DB |||||
1265 CAGTTTGTGTGAGCAGCAACAGCAGATTAAGTCCACTTCCATTTCTGATCACTCG 1324
QY 1365 TACAGGGAACACCGGCTTCTAGCTGAGTACCTCTCTAGCTTCCAGTCCAAACGCGCTGCCA 1424
DB |||||
1325 TACCGGACACCGGCTTCTAGCTGAATACCTCTCTAGCTTCCAGTACCCATCCCG 1384
QY 1425 GGGATCTTCATGTGCAAGACTGAGCGTGCATCCGAAAGGAACTGCGCTGCGACGGCTG 1484
DB |||||
1385 GGGAGTTCAGTGCAGCGGCGCTGATCCGGAAGAGCTGCGCTGTGATGGCTG 1444
QY 1485 GCAGACTGCCCGATTAATAGTATGAGCGTCTTACTGCGGATGCAATGCCACCCACCACTTC 1544
DB |||||
1445 GCGAGCTGCACCGACCCACAGCGATGAGCTCAACTGCAGTTGCGACGCGCGCCACCACTTC 1504
QY 1545 AGTGGAAAAACAGTCTGGAAGCCCTCTTCTGGGTCTGTGAAGTGTCAACAGCTGT 1604
DB |||||
1505 AGTGGAAAAACAGTCTGGAAGCCCTCTTCTGGGTCTGTGAAGTGTCAACAGCTGT 1564
QY 1605 GGGAGCGGAGTGAAGGAGGCGCTGAGCTGTCTGCTGGGAGTTTCAAGTGTCCCAAT 1664
DB |||||
1565 GAGAGCAACAGCGACGAGCGGGTGCAGTTGTCCGCGCCAGACTTTCAGTGTTCAT 1624
QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAACTGTAAATGGGAAGGACAACTGTGGAGATGGTCT 1724
DB |||||
1625 GGGAGTGTCTCCCTCAGAGCCAGAACTGTAAATGGGAAGGACAACTGTGGAGATGGTCT 1684
QY 1725 GAGGAGCTTCATGTGACAGCTGTATGCTGCTCTTGGACCAAAATATACCTACCGCTGC 1784
DB |||||
1685 GAGGAGCTTCATGTGACAGCTGTATGCTGCTCTTGGACCAAAATATACCTACCGCTGC 1744
QY 1785 CAAATGGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGATGGGAAGGACGAGCTGTAGC 1844
DB |||||
1745 CTCATGGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGACGGGAAGGAGGAGTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACTGTGATGTGGGTGCGATCTTTTACCAAAACAGGCTGCG 1904
DB |||||
1805 GACGGCTCAGATGAGAAAGGACTGCGACTGTGGGTGCGGTCACTTACAGAGACAGGCTCGT 1864
QY 1905 GTGGTGGTGGCAACGAAATGCGGACGAGGCGAGTGGCCCTGCGAGTGGCTTCCAGCGC 1964
DB |||||
1865 GTTGTGGGGGACCGATCGGATGAGGCGAGTGGCCCTGCGAGTAAGCTGTGATGCT 1924
QY 1965 CTGGGCGAGGGGCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGTGGTGTCTCTGA 2024
DB |||||
1925 CTGGGCGAGGGGCACTTGTGTGGGGCTCGCTCATCTCTCTCCCACTGGCTGTCTCTGCC 1984
QY 2025 GCTCATGTCTTCAGATGACAAAAATTTCAAGTACTCAGATCAACAGATGTGAGCGGC 2084
DB |||||
1985 GCACACTGTCTACATCGATGACAGAGATTCAGGTACTCAGACCCCAACGCACTGGAGCGGC 2044
QY 2085 TTCTCTGGGTCTGTGACAGGAGGAGCGAGTGGCTGCTGGGGTGCAGAGCTGAGGCTC 2144
DB |||||
2045 TTCTCTGGGTCTGTGACAGGAGGAGCGAGCGCCCTGGGGTGCAGAGCGAGGCTC 2104
QY 2145 AAAAGTATCATCAACCCCTCTCTTCAATGATTTTCACTTGCATATGACATCGCTTGT 2204
DB |||||
2105 AAGGATCATCTCCACCCCTCTTCAATGATTTTCACTTGCATATGACATCGGCTG 2164
QY 2205 CTGGAGCTGGAGAGTGTGGTGGAGTACAGCACCGTGTGCGGCCCACTTGTGCTGCTGAT 2264

Db	2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGTCGGGCCCATCTGCTTCGCGGAC	2224
QY	2265 GCTTACCATTCTTTCCCTGTCTGCAAGGCCATCTGGGTACAGGCTTGGGGCACAAAA	2324
Db	2225 GCCTCCCATGCTTCCCTGCGCGCNAAGGCCATCTGGGTACGGGCTTGGGACACACCCAG	2284
QY	2325 GAGGAGGTAACGGAGCGCTGATCCTTGCAAGAGGTGAGATCCGTGTCAACAACGAGCC	2384
Db	2285 TATGGAGGCATCTGGCGCGCTGATCCTTGAAAAGGCTGAGATCCGCGCTCATCAACGAGACC	2344
QY	2385 ACCTGTGAGACCTCATGSCGACGACATCACCCACGAATGATGTGTGGGTTTCCTTC	2444
Db	2345 ACCTGCGAAGACTCTCTGCGCGACGATCACGCGCGCATGATGTGCGGTGCTTCCTC	2404
QY	2445 AGTGGGGTGTGAACTCCTGCCAAGGTTGACTCTGTGTGGCCCCCTTGTCAAAGCGCGAGA	2504
Db	2405 AGCGCGCGCTGGACTCCTGCCAAGGTGATTCGGGGGACCCCCTGTCCAGCGTGGAGCG	2464
QY	2505 GATGGGCGAATGTTCCAGCTGTGTGGTGAGCTGGGCTGAAGGCTGCGCTCAGAGAAC	2564
Db	2465 GATGGGCGGATCTTCCAGGCGGTGTGGTGAGCTTGGGGAGACGGCTCGCTCAGAGAAC	2524
QY	2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTTCGGGACTTGGATCAAAGACACACTGGG	2624
Db	2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTTGGATCAAAGAGAACCTGGG	2584
QY	2625 GTATAGCAGCATGGAACAGACGCGGACCAAAACACCCACAGGGATGCCGACATGCACA	2684
Db	2585 GTATAGGGGCGGG--GCCACCTCAAATGTGTATACCTGCGGGGCCCACCATGTCACCC	2641
QY	2685 CTTGGATACAGGAGAGAACACTGACACATTTATGCTGTGGCTCTCCCCCCCCCAACACA	2744
Db	2642 CCAGTGTGCAG-CTTGACGCTTGGAGCTGGACCGCTGACTGCACCGAGCGCCC-CCAG	2699
QY	2745 ACCAGACTGTGAACCTGCATCTCTTAGGACTCAGAGT	2780
Db	2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT	2735

Search completed: February 21, 2004, 05:28:26
Job time : 11176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:03:32 ; Search time 795 Seconds
(without alignments)

10546.484 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catggtagacggctgccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3106	100.0	3106	AAD37039	Epithin gene. Uni
2	1889.8	60.8	3149	21	Human matrixase c
3	1884.8	60.7	3142	22	Human membrane-tyr
4	1883.2	60.6	3147	20	Tumour antigen der
5	1883.2	60.6	3147	22	Human membrane-tyr
6	1883.2	60.6	3147	22	Human membrane-tyr
7	1883.2	60.6	3147	22	Human membrane-tyr
8	1883.2	60.6	3147	22	Human TADG-15 codi
c	1883.2	60.6	3147	22	Human TADG-15 anti

9	1883.2	60.6	3147	24	AAL53444	Type II transmembr
10	1883.2	60.6	3147	24	AAL53445	Type II transmembr
11	1883.2	60.6	3147	25	ABZ58500	Transmembrane seri
12	1883.2	60.6	3147	25	ABZ58501	Transmembrane seri
13	1883.2	60.6	3147	25	ABZ22450	Human membrane-tyr
14	1883.2	60.6	3147	25	ABZ22451	Human MTSPI protea
15	1883.2	60.6	3147	25	ABZ47180	Human membrane-tyr
16	1883.2	60.6	3147	25	ABZ47181	Human membrane-tyr
17	1883.2	60.6	3147	25	ABZ47225	Human membrane-tyr
18	1883.2	60.6	3152	22	AAH57431	Human intestine ce
19	1872.8	60.3	3159	21	AAH37657	Human peptidase, H
20	1793.8	57.8	3115	22	ABA08672	Human membrane-tyr
c	1793.8	57.8	3115	22	AAH95659	Human protein enco
22	1762.2	56.7	2955	21	AAA88492	Human matrixase (
23	1559.2	50.2	2756	21	AAH77957	Human cancer assoc
24	1532.4	49.3	3112	24	ABH76526	CDNA encoding huma
25	1530.8	49.3	2900	22	AAH23602	Human SNC-19 codin
26	1498.8	48.3	3413	23	AAH84260	Human cDNA differe
27	1498.8	48.3	3413	23	AAH85629	DNA encoding novel
28	701.2	22.6	1553	24	ABK30240	Human G-protein-co
29	556.4	17.9	726	24	AAH36928	Human matrixase o
30	551.8	17.8	723	18	AAH79128	Human serine prote
31	379.4	12.2	2086	23	AAH85628	DNA encoding novel
32	367	11.8	1458	23	AAH85625	DNA encoding novel
33	306.8	9.9	434	24	ABK39442	DNA encoding lung
34	306.8	9.9	434	25	ACA11771	Human lung cancer
35	306.8	9.9	434	25	ACA02957	Lung cancer therap
36	302.6	9.7	429	24	ABL82332	Human ovarian canc
37	279.2	9.0	387	23	AAH85626	DNA encoding novel
38	265	8.5	591	23	AAH58827	CDNA #1503 encodin
39	246.8	7.9	393	25	ABH40789	Bovine EST associa
40	242.4	7.8	2801	24	ABQ61196	Human PRO618 encod
41	240.8	7.8	2672	22	ABQ13117	Human membrane-tyr
42	240.8	7.8	3104	22	AAH13116	Human membrane-tyr
43	239.2	7.7	1327	20	AAZ34034	Human EST DNA35597
44	239.2	7.7	1327	21	AAH78495	Human EST DNA35597
45	239.2	7.7	1327	21	AAH58237	Human EST (express

ALIGNMENTS

RESULT 1
AAD37039
ID AAD37039 standard; DNA; 3106 BP.
AC AAD37039;
XX
XX
XX 21-AUG-2002 (first entry)
XX Epithin gene.
XX Transgenic; transgenic animal; pharmacological therapy; gene therapy;
XX phenotype modulation; genetic disease; epithin; gene; ds.
XX Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 63...2630
XX /*tag= a
XX /*product= "Epithin protein"
XX /*tag= b
XX /*note= "Sequence flanking Neo insert in
XX targeting construct"
XX /*tag= c
XX /*note= "Sequence deleted in targeting construct"
XX /*tag= d
XX /*note= "Sequence flanking Neo insert in
XX targeting construct"

1283 CGTGGTCAACAGCAACAGCAACAGATCACAGTTCGCTTCCACTCAGATCAGTCTCAAC 1342
1370 GGACACGGGTTCTAGCTAGTACCTCTCTAGACTCCAAAGACCGTCCAGGAT 1429
1343 CGACACGGCTCTTAGCTGAATACCTCTCTAGACTCCAGTACCCATCCCGGGGCA 1402
1430 GTTCATGTGCAAGACTGACGGTGCATCCGAAGGAACTGCGCTGCAGCGGCTGGGAGA 1489
1403 GTTCACGTGCGGCAACGGGGCGGTGTATCCGAAGAGAGCTGCGCTGTGATGGCTGGCGCA 1462
1490 CTGCGCGGATTAATGATGAGGTTACTGCGGATGCAATGCCACCCACAGTTCAGGTG 1549
1463 CTGACCGACACACGAGATGAGCTCAACTGCAAGTTGGACGCCCGCCACCAATTCAGGTG 1522
1550 CAAAAACCAAGTCTGCAAGGCCCTCTCTTGGGTCTGTGACAGTGTCAACGACTGTGGGGA 1609
1523 CAAGACAAAGTTCTGCAAGCCCTCTCTTGGGTCTGTGACAGTGTGAACGACTGCGGAGA 1582
1610 CGGAAGTGACAGGAGGCTGCAGCTGTCTGCTGGAGTTTCAAGTGTTCCTCAATGGGAA 1669
1583 CAACAGCGACGAGCGGGTGCAGTGTGTCGGGCCACAGACCTTCAGGTGTTCCTCAATGGGAA 1642
1670 GTGCTCTCCTCAGACCCAGAGTGTATGGGAAGGACAACTGTGGAGATGGGTCTGACGA 1729
1643 GTGCTCTCGAAAGCCAGCAGTGCATGCAATGGGAAGGACGACTGTGGGACCGGTCCGACGA 1702
1730 GGCCTTCATGTGACAGCGTGAATGTCGTCTTGTGACCAAAATATACCTACCGCTGCCAAAA 1789
1703 GGCCTCTGCCCCAAGGTGAACGTGCTCACTTGTACCAACACACCTACCGCTGCCCTCAA 1762
1790 TGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGATGGGAAGACGGAATGTAGCGATGG 1849
1763 TGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGACGGGAAGGAGACTGTAGCGACGG 1822
1850 CTCGATGAGAAAACCTGCACTGTGGGCTGCCATCTCTTTACCAACAGAGCTGCCGTGT 1909
1823 CTCAGTGAAGAGACTGCGCACTGTGGGCTGCGGTCACTTCAACGAGACAGGCTGTGTGT 1882
1910 TGTGGCAGCAATGCGGACGAGGCGAGTGGCCCTGCGCAGTGAAGCTCCACAGCCCTGGG 1969
1883 TGGGGGACCGATGCGGATGAGGCGAGTGGCCCTGCGCAGTGAAGCTGTAGCTGTGGG 1942
1970 CAGGGGCACTGTGTGGGCGCTGCTCACTCTCTCTGACTGGCTGGTCTCTGACGCTCA 2029
1943 CAGGGGCACTGTGGGCTGCTTCCCTCACTCTCTCCCACTGGCTGTCTCTGCGGCA 2002
2030 TTGCTTTCAAGATGACAAAATTTCAAGTACTCAGACTACAGATGTGGACGGCTTCTCT 2089
2003 CTGCTACATCATGACAGAGGATTCAGGTACTCAGACCCCAACGAGTGGACGGCTTCTCT 2062
2090 GGTCTCTGACAGCAAGCGCAGTGGCTCTTGGGCTGACGAGCTGAAGCTCAAAACG 2149
2063 GGGCTTGACAGCAAGCGCAGCGCCCTGGGCTGACGAGCGCAGGCTCAAGCG 2122
2150 TATCATCAACCACTCTCTTCAATGATTTCACTTCACTATGACATGCCCTGTGTGGA 2209
2123 CATCATCTCCCACTCTTCAATGATTTCACTTCACTATGACATGCCCTGTGTGGA 2182
2210 GCTGGAGAGCTCGGTGAGTACAGCAGCTGTCGCGCCCTCATCTGCTGCTGATGCTAC 2269
2183 GCTGGAGAAACCGGACAGTACAGTCCATGTTGCGGCCCTCATCTGCTGCGGACGCTC 2242
2270 CCAATGCTTCCCTGCTGGCAAGGCACTCTGGGTCAACAGGCTGGGGGCAACAAAGAGGG 2329
2243 CCAATGCTTCCCTGCGGCAAGGCACTCTGGGTCAACAGGCTGGGGGCAACCCAGATGG 2302
2330 AGGTACCGGAGCGCTGATCTCTGAGAGGAGTGAATCCGTGTCTATCAACAGACCACTG 2389
2303 AGGCACTTGGGCGCTGATCTCTGAAAAGGAGTGAATCCGGGTCTATCAACAGACCACTG 2362
2390 TGAGGACCTCATGCGGACGAGATCAACCCACGATGATGTGTGGGTTCCTCAGTGG 2449

2363 CGAAGACCTCTGCGCAGCAGATCAAGCGCGCATGATGTGCTTCTCCTCAGCGG 2422
2450 GGGTGTGGACTCTCTGCGCAGGGTACTCTGTGGCCCTTGTCAAGCGCGGAGAAAGATGG 2509
2423 CGGGTGGACTCTCTGCGCAGGGTGAATCCGGGGGACCCCTGTCTCCAGCGTGGAGCGGATGG 2482
2510 GCGAATCTTCCAGGCTGCTGTGGTGAAGTGGGGTGAAGGCTGCGCTCAGAGGAAACAAGCC 2569
2483 GCGGATCTTCCAGGCGCGTGTGGTGAAGTGGGAGACGGTGTGCTCAGAGAACAGCC 2542
2570 AGGCTGTACAAAGGCTCTCTGTAGTTCGGGACTGGATCAAGAGACACTGGGGTATA 2629
2543 AGGCTGTACAAAGGCTCTCTGTGTTTCGGGACTGGATCAAGAGAACACTGGGGTATA 2602
2630 GCAGATGGACAGACAGCGCCGACCAACACCCACAGGGATGCCGACATGCACACCTGG 2689
2603 GGGCGCGG---GCCACCCAAATGTGTACCTGTGGGGGACCCCATGCTCCACCCCACT 2659
2690 ATACAGAGAGGAACACTGACGACATTTATGCTGTGGCTTCCCGCCCCCAACCAACCA 2749
2660 GTGACG-CCTGCGAGCTGGAGACTGGACCGCTGACTGTGACACAGCGCCC-CCAGAACATA 2717
2750 GACTGTGAACCTGCATCTTTAGGACTCAGAT 2780
2718 CACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2748

RESULT 3

AAF28099 standard; cDNA; 3142 BP.

AC AAF28099;

XX 06-JUN-2001 (first entry)

XX Human membrane-type serine protease MT-SPI coding sequence.

XX Human; membrane-type serine protease; MT-SPI; cancer; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
CDS 37..2602

FT /*tag= a

FT /product= "MT-SPI"

XX WO200123524-A2.

XX 05-APR-2001.

XX 02-OCT-2000; 2000WO-US27250.

XX 30-SEP-1999; 99US-0410362.

XX (REGC) UNIV CALIFORNIA.

XX Craik CS, Takeuchi T, Shuman M;

XX WPI; 2001-245002/25.

XX P-PSDB; AAB35465.

XX New nucleic acid encoding a membrane type serine protease, useful for the diagnosis, prognosis and treatment of cancer, particularly metastatic cancers -

XX Claim 4; Fig 1; 102pp; English.

XX The present invention provides the protein and coding sequences for the novel human membrane-type serine protease MT-SPI. Increased expression of this protein is associated with cancer, and so the sequences can be used in cancer diagnosis and the identification of treatments. The present sequence is the MT-SPI coding sequence.

XX

SQ	Sequence	3142 BP; 635 A; 958 C; 965 G; 584 T; 0 other;	
	Query Match	60.7%; Score 1884.8; DB 22; Length 3142;	
	Best Local Similarity	82.8%; Pred. No. 0;	
	Matches 2164; Conservative	0; Mismatches 447; Indels 1; Gaps 1;	
QY	34	CTGAGACCGGCGATCGGACCGCCAAACCAATGGGTAGCAATCGGGCCCGCAAGCCCGGAG	93
DB	9	CTGAGACCGCGCAGCGG-CCTCGGGACCACTGGGAGCGATCGGGCCCGCAAGGGCGGAG	67
QY	94	GGGGCTCTCAGGACTTCGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAAATGAATG	153
DB	68	GGGGCCCGAAGGACTTCGGCGCGGGACTCAAGTACAACTCCCGGCAAGGAAAGTGAATG	127
QY	154	GCTTTGAGGAGGGTGTGAGTTCCTGCTCGCAACAAATGCCAAGAAAGTGGAGAAGCGAG	213
DB	128	GCTTTGAGGAAGCGGTGGAGTTCCTGCCAGTCAACAAAGTCAAGAAAGTGAATG	187
QY	214	GCCCCAGGCGCTGGGTGGTCTGGTGGCAGTGTCTTTCAGCTTCCCTTGGCTCTCCCTCA	273
DB	188	GCCCCGGGCGCTGGGTGGTCTGGCAGCGGTGTGATCGGCGCTCTTGGTCTTGCTGG	247
QY	274	TGGCTGGCTTGTGCTGTGGCACTTTCANATATCGGAATGTGCGGGTTCAAAAAGTCTTCA	333
DB	248	GGATCGGCTTCTGCTGTGGCACTTTCAGTACCGGGACGTCGCTGTGCAAGAGTCTTCA	307
QY	334	ATGGCCATCTGAGGATCAAAATGAGATCTTCTTGGATGGTATGAGAACTCCACCTCCA	393
DB	308	ATGGCTACATGAGGATCAAAATGAGAAATTTTGGATGCCCTACGAGAACTCCAACTCCA	367
QY	394	CAGAGTTTATCAGCTGGCGCAGCAGGAGTGAAGGAGGCGCTGAAGTGTCTGTAAATGAAG	453
DB	368	CTGAGTTTGTAAAGCTGGCGCAGCAAGGTGAAGACGCGCTGAAGTGTCTGTACAGCGGAG	427
QY	454	TCCTGTCTGGTCCCTACCAAGAGTGGCTGTAACTGCTTCAAGTGGGCGAGTG	513
DB	428	TCCATTCTTGGGCCCTTACCAAGAGTGGCTGTGACGGCTTTCAGGAGGGCAGCG	487
QY	514	TCATCGCTTACTACTGGTCAGAGTTTCAGATCCCCCACACCTGGCAGAGAGAGTTGATC	573
DB	488	TCATCGCTTACTACTGGTCTGAGTTTCAGATCCCGCAGCACCTGGTGGAGAGGCGCGAGC	547
QY	574	GCGCATGGCTGTGAGAGGAGTTGAACTTGCACACCCGAGCAGCGGCACTGAATTCCT	633
DB	548	GCCTCATGGCCGAGAGCGGTAGTCACTGTGCCCGCGCGCGCTCCCTGAAGTCTCT	607
QY	634	TCGTGTAAATCTGTGTGGGCTTCCCAATGACCCCAAGATGCTGCAGAGGACTCAGG	693
DB	608	TTGTGTCACTCAGTGGTGGCTTTCCCCACGACTTCCAAAACAGTACAGAGNCCAGG	667
QY	694	ACAACAGCTGCAGTTTGGCCCTGCATGGCCATGGTGACAGTGAACCGTTCACTACCC	753
DB	668	ACAACAGCTGCAGTTTGGCTTGCACGCCCCGCGGTGTGAGGCTGATCGGCTTCAACACGC	727
QY	754	CTGGCTTCCCCAAGTCCCTACCCGCGCATGCCGCTGCCAGTGGGTCTCTCGGGGGGG	813
DB	728	CCGCTTCCCTGACAGCCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCGCTCGGGGGG	787
QY	814	ACCGGACTCTGTGTGAGGCTTCACTTCCGAAGCTTTGATGTGCTCCCTGTGATGAGC	873
DB	788	ACCGGACTCAGTGTGAGGCTTCACTTCCGAGCTTTGACCTTGTGCTCTTCGCGAGCAGC	847
QY	874	ATGGCAGTGAAGTGTCAACGGTGTATGAGTGTAGCCCACTGGAACCCCAAGCTGTGG	933
DB	848	GCGGACGAGCACTGGTGAACGCTGTACAAACCCCTGAGCCCCATGGAGCCCAAGCCCTGG	907
QY	934	TGCGGCTGTGTGGCACTTCTCACCTCCTTACAACTGACTTTGCTCTCTCTCCCAAGACG	993
DB	908	TGCAGTTGTGTGGCACTTACCTCTCTTCTACAACTGACCTTCCACTCTCTCCAGAACG	967
QY	994	TCTTCTCTGTGACGCTGATTAACAAATACTGACCGGCGACATCTCTGGCTTTGAGGCCACTT	1053
DB	968	TCTGCTCATCACTGATTAACAAACTGACGAGCGCGGCACTCCGCGTTTGTGAGGCCACCT	1027

2134 AGCTAGAGCTCAACAGCTATCATACCCACACCTCTCTCAATGATTTACCTTCGACTATG 2193
 2108 AGCGAGGCTCAAGCGCATCATCTCCACACCTCTCTCAATGATTTACCTTCGACTATG 2167
 2194 ACATCGCTTGTGGAGCTGGAGTCTGGTGGAGTACAGACAGCTGCTGGGCCCATCT 2253
 2168 ACATCGGCTGCTGGAGCTGGAGAAACCGGACAGTACAGTCCATGCTGGGCCCATCT 2227
 2254 GCCTGCTGATGTACCATGCTCTCTCTCTGTCGCAAGGCTCATCTGGGTACAGGCTGG 2313
 2228 GCCTGCGGAGCTCTCCATGCTCTCTCTGCGCAAGGCTCATCTGGGTACAGGCTGG 2287
 2314 GGCACAAAGAGGAGGAGTACCGGAGCTGATCTCTGCAAGAGGCTGAGATCCGTCTCA 2373
 2288 GACACACCCAGTATGGAGCACTGGCGGCTGATCTCTGCAAGAGGCTGAGATCCGCTCA 2347
 2374 TCACACAGACCACTGTGAGGACCTCATGCGGAGGAGATCACCCACAGATGATGTGTG 2433
 2348 TCACACAGACCACTGTGAGGACCTCTCTGCGGAGGAGATCACCGCGCATGATGTGTG 2407
 2434 TGGGTTTCTCAGTGGGCTGTGACTCTCTGCGGAGGCTGATCTGTGTGGCCCTTGTCAA 2493
 2408 TGGGCTTCTCAGCGGCGGCTGACTCTCTGCGGAGGCTGATCTGTGTGGCCCTTGTCAA 2467
 2494 GCGCGGAGAAAGATGGCGCAATGTTCCAGGCTGGTGTGGTGGAGCTGGGGTCAAGCTGG 2553
 2468 GCGTGGAGGCGGATGGCGGATCTTCAGGCGGCTGTGGTGGAGCTGGGGAGAGCGCTGG 2527
 2554 CTGAGGAGACACCGAGGCTGTACACAGGCTCTCTGTAGTTCGGGAGCTGATCAAG 2613
 2528 CTGAGGAGACACCGAGGCTGTACACAGGCTCTCTGTAGTTCGGGAGCTGATCAAG 2587
 2614 AGCACACTGGGCTATAGCAGCATGGACAGACA 2645
 2588 AGAACCTGGGCTATAGGGCGGGGACACCA 2619

RESULT 4

AAx87815

ID AAX87815 standard; cDNA; 3147 BP.

XX

AC AAX87815;

XX

XX 09-NOV-1999 (first entry)

XX Tumour antigen derived gene-15 (TADG-15) cDNA.

XX

XX Tumour antigen derived gene-15; TADG-15; serine protease; human;

XX breast cancer; ovary cancer; carcinoma; diagnosis; ss.

XX

XX Homo sapiens.

XX

XX Key Location/Qualifiers

XX CDS 23..2590

XX /*tag= a

XX

XX WO9942120-A1.

XX

XX 26-AUG-1999.

XX

XX 18-FEB-1999; 99WO-US03436.

XX

XX 20-FEB-1998; 98US-0027337.

XX

XX (UYAR-) UNIV ARKANSAS.

XX

XX O'Brien TJ, Tanimoto H;

XX

XX WPI; 1999-527418/44.

XX

XX P-PSDB; AAY06671.

XX

XX A new extracellular serine protease for diagnosis of neoplastic

XX disease

XX

XX

XX

PS

XX

XX Claim 2; Fig 9; 71pp; English.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match

Best Local Similarity

Matches 2223; Conservative

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Db	725	GACAGCCCTTACCCCGCTCTATGCCCCTGCGCAGTGGGCCTCTGCGGGGGAGACGCCGACTCA	784
Qy	825	GTGCTGAGCCTCACTTCCGAAGCTTTTGATGTGCTCTCCCTGTGATGAGCATGGCAGTGCAC	884
Db	785	GTGCTGAGCCTCACTTCCGAGCTTTTGACTTTCGTCGTCTGCGACGAGCGCGCAGCGAC	844
Qy	885	CTGCTACCGTGATGATAGCCTGAGCCCATGGAAACCCACGCTGTGGTGTGGCGCTGTGT	944
Db	845	CTGGTGACGGTGTTACAACAACCTGAGCCCCATGAGCCCCACGCGCTTGGTGCAGTTGTGT	904
Qy	945	GGCACCTTCTCACCCCTCTAAACCTGACTTTCTCTCTCTCCAGAAAGCTTCTTCTTGTCT	1004
Db	905	GGCACCTTACCTCTCCCTCTCAACCTGACCTTCCACTCTCTCCAGAAAGCTCTCTGCTCATC	964
Qy	1005	ACGCTGATACCAATACGACCGGCGACATCTCTGGCTTTGAGGCGCATTTCTTCTCAGCTG	1064
Db	965	ACACTGATACCAACACTGAGCGCGGCATCTCCCGCTTTGAGGCCACCTTCTTCTCAGCTG	1024
Qy	1065	CCCAAGATGACGAGCTGTGGCGGCTTTTTCAGTGTGACACCCAAAGGACATTTAGCAGCCCC	1124
Db	1025	CCTAGGATGAGAGCTGTGGAGCGCTTACGTAAAGCCCAAGGACATTTCAACAGCCCC	1084
Qy	1125	TACTATCCAGGCCACTACCGGCCCAACATCAACTGACATGCAATGGAATATCAAGGTGCCAAC	1184
Db	1085	TACTTACCCAGGCCACTACCCACCCAAATTTGACTGCACATGGAACATGAGGTGCCAAC	1144
Qy	1185	AACCGGAACTGTAAGGTGCGCTTCAAACCTCTTCTATCTGTGTGACCCCAACGATACCAGTG	1244
Db	1145	AACCGACATGTGAAGTGAAGTTCAAATCTTCTACTGTGTGAGCCCGGCTGCTGG	1204
Qy	1245	GGCTCTGTGACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGTCC	1304
Db	1205	GGCACTTGCCTCCAGGACTACGTGGAGATCAATGGGGAGAGAACTCTCGGAGAGAGTCC	1264
Qy	1305	CAGTTTCTGTGTGAGCAGCAACAGCAGCAAGATPACAGTCCACTTCCATCTGATCACTCG	1364
Db	1265	CAGTTCGTCTCACCAGCAACGACAGATCACAGTTCGCTTCCACTCAGATCACTGTC	1324
Qy	1365	TACACGACACCCGGTTCCTAGCTGAGTACTCTCTCTACGACTCCAAAGCACCCGCTGCCCA	1424
Db	1325	TACACGACACCCGGTTCCTTAGCTGAATACCTCTCTTACGACTCCAGTGACCCATGCCCG	1384
Qy	1425	GGGATGTTCTATGTGCAAGACTGACAGCGTGCATCCGAAGAAGAACTGCGCTCGCAGCGCTGG	1484
Db	1385	GGGCAGTTCAGTGTGCGCAGCGGGCGGTGTATCCGGAAGAGCTGCGCTGTGATGCGCTGG	1444
Qy	1485	GCAGACTGCCGGATTTATGTTGATGAGCGTTACTGCCGATGCAATGCCACCCACCACTTC	1544
Db	1445	GCCGACTGCACCGACCAACAGCATGAGTCAACTGTCAGTTCGCGCGCGCCACCACTTC	1504
Qy	1545	ACGTGCAAAACAGTTCTCGAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT	1604
Db	1505	ACGTGCAAGAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGCAAGTGTGAACGACTGC	1564
Qy	1605	GGGACCGGAAGTCAACGAGGAGGGCTGCAGCTGCTCTGCTGGGAGTTTCAAGTGTCCAAT	1664
Db	1565	GGAGACAACAGCGACGAGGAGGGTGCAGTTGTTCGGCCACGACTTTCAGTGTTCAT	1624
Qy	1665	GGGAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGCAACTGTGGAGATGGGTCT	1724
Db	1625	GGGAAGTGCCTCTCGAAAGCCAGCAGTGCAAATGGGAAGCAGCACTGTGGGACGGGTCC	1684
Qy	1725	GACGAGCTTCATGTGACAGCGTGAATGTCGTCTCTTGACCAAAATATACCTACCCTGC	1784
Db	1685	GACGAGGCTCTCTGCCCAAGTGAAGTGTCACTTGTACCAAAACACCTACCCTGTC	1744
Qy	1785	CAAAATGGCCTCTGCTGTGACCAAGGCAACCCCTGAGTGTGATGGGAAGCAGCACTGTAGC	1844
Db	1745	CTCAATGGGCTCTCTTCTGACCAAGGCAACCCCTGAGTGTGACGGGAAGGAGCACTGTAGC	1804
Qy	1845	GATGGCTCCGATGAGAAATACTGTGACTGTGGCTTGGATCTCTTTACAAACAGGCTCGC	1904
Db	1805	GACGGCTCAGATGAGAGGACTGCAGTGTGGCTGGGTCGCTCATTCACGACAGGAGTCTGT	1864

Qy	1905	GTGGTTGTGTGGCAAGAAATGCGAGACGAGGCGAGTGGCCCTCTGGCAGGTGAGCCCTCCACGCC	1964
Db	1865	GTTTGTGGGGGCAACGATGCGGATGAGAGGCGAGTGGCCCTCTGGCAGGTAAAGCCTGCAATGCT	1924
Qy	1965	CTGGGCGAGGGCCACATTGTGTGGGGCTCGCTCATCTCTCTGACTGGTGGTCTCTTGCA	2024
Db	1925	CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCACTGGCTGGTCTCTGCC	1984
Qy	2025	GCTCAATGCTTTTCAGGATGACAAAAATTTCAAGTACTCAGACTPACAGATGTGGACGGCC	2084
Db	1985	GCACACTGCTTACATCATGATCAGACAGAGATTTCAGGTACTCAGACCCCAAGCTGGAGCGGCC	2044
Qy	2085	TTCTCGGTCTGCTGGACAGACGACGAGTGCCTCTGGGGTGCAGCAGCTGAGAGCTC	2144
Db	2045	TTCTCGGTCTGACAGACAGACGACGAGCGCCCTTGGGGTGCAGAGCGCAGCTC	2104
Qy	2145	AAACGTATCATCACCAACCTTCCTTCAATGATTTTCACTTCGACTATGACATCGCCTTG	2204
Db	2105	AAGCGCATATCTCCCAACCCCTTCTTCAATGACTTTCACCTTCGACTATGACATCGCGCTG	2164
Qy	2205	CTGGAGCTGGAGAGTCTGGTGGAGTACAGCACGCTGCTGGGCCCACTCTGCCCTGCTGAT	2264
Db	2165	CTGGAGCTGGAGAAACCGCAGAGTACAGCTCCANGTTCGGCCCACTCTGCCCTGGCGAC	2224
Qy	2265	GCTTACCATGCTTTCCTCTGTGCGCAAGGCCATCTGGGTCAACAGGCTGGGGCACACAAA	2324
Db	2225	GCCTCCCATGCTTCTCTGCGGCAAGGCCNCTGGGTCAAGGCTGGGGACACACCCAG	2284
Qy	2325	GAGGAGGTACCGGAGCGCTGATCTCTGCGAAGGGTGGAGTCCGTGTCTCATCAACACAGACC	2384
Db	2285	TATGGAGGCACTGGCGCGTGTACTCTTGCAAAAGGGTGAGATCCGCGTCTATCAACACAGACC	2344
Qy	2385	ACCTGTGAGGACCTCATGCGCGACGAGATCACCCACGAAATGATGTGTGGGTTCCTC	2444
Db	2345	ACCTGCGAGAACCTCTGCGCGACGATCTACGCCGCGCATGATGTGCGTGGGTTCCTC	2404
Qy	2445	AGTGGGGTGTGACTCTCTGCCAGGTGACTCTGTGTGGCCCTTTGTCAAGCGCGAGAAA	2504
Db	2405	AGCGGCGCGCTGAGCTCTGCCAGGCTGATTCGCGGGGACCCCTGTGCCAGCGTGGAGGGC	2464
Qy	2505	GATGGCGAATGTTCCAGCTGTGTGTGTGAGCTGGGTTGAGGCTGCCTCAGAGGAC	2564
Db	2465	GATGGCGGATCTTCCAGGCGCGTGTGTGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC	2524
Qy	2565	AAAGCCAGGCGGTATACAAAGGCTCCCTGTAGTTTCGGGACTGGATCAAGAGCACACTGGG	2624
Db	2525	AAAGCCAGGCGGTATACAAAGGCTCCCTGTGTTCGGGACTGGATCAAGAGAACACTGGG	2584
Qy	2625	GTATAGCAGCATGGACAGACAGCGGACCAACACCCACAGGATGCCGACATGCACA	2684
Db	2585	GTATAGGGGCGGG---GCCACCCAATGTGTACACTCGGGGCGCACCATCTGCTCACCC	2641
Qy	2685	CTTGAGTACAGGAGAGGAACACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACA	2744
Db	2642	CCAGTGTGCACG---CTTGAGGCTGGAGACTGGACCGCTGACTGCACACAGCGCCC---CCAGA	2699
Qy	2745	ACCCAGACTGTGAACTGCAATCTCTTAGGACTCAGAGT	2780
Db	2700	ACATACACTGTGAATCAATCTCCAGGGTCCAAAT	2735

RESULT 5

RESUL 3
AAD13113

AAD13113
ID AAD13113 standard; DNA; 3147 BP.

ID AAD13113 :
XX

XX
AC
BADI3113.

AC AAD13113;
vv

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

DT 16-OCT-2001 (first entry)

XX

DE Human membrane-type serine protease (MTSP) 1 DNA.

MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic; lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy; matriptase; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 23..2590

XX FT /*tag= a

XX FT /product= "Human transmembrane serine protease (MTSP) 1"

XX PN WO200157194-A2.

XX 09-AUG-2001.

XX 02-FEB-2001; 2001WO-US03471.

XX 03-FEB-2000; 2000US-0179982.

XX 18-FEB-2000; 2000US-0183542.

XX 22-JUN-2000; 2000US-0213124.

XX 26-JUL-2000; 2000US-0220970.

XX 08-SEP-2000; 2000US-0657986.

XX 22-SEP-2000; 2000US-0234840.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO, Yeh J;

XX WPI; 2001-488877/53.

XX P-PSDB; AAE06930.

XX Novel single chain polypeptide comprising protease domain of type-II membrane-type serine protease or its catalytically active portion useful for treating and preventing cancer and tumor -

XX Claim 14; Page 191-195; 256pp; English.

XX The invention relates to transmembrane serine proteases and their corresponding nucleotides and the protease domain of a type-II membrane-type serine protease (MTSP). MTSP is useful for identifying compounds that modulate or inhibit its proteolytic activity and for formulating a medicament for treating neoplastic disease. MTSP and its corresponding nucleotides are useful in preventing or treating tumours or cancers such as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP is useful as a diagnostic marker for tumour development, growth and/or progression and as immunogens to generate antibodies that specifically bind to it. MTSP DNA is useful in a yeast two-hybrid system and in gene therapy. The present sequence is a DNA encoding human MTSP1 protein (also called matriptase).

XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

XX Query Match 60.6%; Score 1883.2; DB 22; Length 3147;

XX Best Local Similarity 81.2%; Pred. No. 0;

XX Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGGGGCTCTCAG 104

DB 5 GAGCGGCTCGGGGTACATGGGAGGATCGGGCCCGCAAGGGCGGGGGCCGAAG 64

QY 105 GACTTCGGCGGGGACTCAAGTCAAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAG 164

DB 65 GACTTCGGCGGGGACTCAAGTCAAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAA 124

QY 165 GGTGTGGAGTTCCTGCTCGCAACATGCGCAAGAAAGTGGAGAGCGAGCGCCCGCGGC 224

DB 125 GGCCTGGAGTTCTGCGCCAGTCAACACGCTCAAGAAAGTGGAAAAGCATGGCCCGGGCGC 184

QY 225 TGGGTGGTGGTGGGAGTCTGTTCAGTCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 284

DB 185 TGGGTGGTGGTGGGAGTCTGTTCAGTCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 244

QY 285 CTGGTGGGCACTTCCATTATCGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

DB 245 CTGGTGGGCACTTGGCAGTACCGGACGTGCGTGTCCGAAAGTCTTCAATGGCTACATG 304

QY 345 AGGATCAAAATGAGATCTTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

DB 305 AGGATCAAAATGAGAAATTTTGTGGATGCGCTACGAGAACTCCAACTCCACTGAGTTTGTGA 364

QY 405 AGCTGCGGACGAGGTGAAGGAGGCGCTGAAGTCTGTACATGAGTCCCTCTCTCTCTCTCT 464

DB 365 AGCTGCGGACGAGGTGAAGGAGGCGCTGAAGTCTGTACATGAGTCCCTCTCTCTCTCTCT 424

QY 465 GGTCTCTTACCACAAGAGTGGCTGTAACTGCGCTTCACTGAGGCGAGTGTCTCATCGCTAC 524

DB 425 GCGCTCTTACCACAAGAGTGGCTGTGACGCGCTTTCAGCGAGGCGAGCTCATCGCTTAC 484

QY 525 TACTGTCTCAGGTTTACGATCCCGCCCACTGCGCAGAAAGGTTGATCGCCCATGGCT 584

DB 485 TACTGTCTCAGGTTTACGATCCCGCCCACTGCGCAGAAAGGTTGATCGCCCATGGCT 544

QY 585 GTGGAGCGAGTTGTAACATTTGCCACCGGAGCAGCGGACCTGAAATCTTCTGCTCTAACA 644

DB 545 GAGGAGCGGTTAGTCTATGCTGCGCGCGCGCGCTCCCTGAACTCTTTGTGGTCACTC 604

QY 645 TCTGTGTGGTGGCTTTCCCAATTTGACCCCGAGAAATGCTGCAGAGGAGTCTCAGGACAAACAGCTGC 704

DB 605 TCAGTGGTGGCTTTCCCGACGAGTCTCCAAACAGTACAGAGGAGCCAGGACAAACAGCTGC 664

QY 705 AGTTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 764

DB 665 AGCTTTGGCTTGCACGCGCGCGGTGTGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 724

QY 765 AACAGTCTCTTACCGGCGAGTGTGCGCTGCGCTGAGTGGTCTGCGGGGGGAGCGGAGTCT 824

DB 725 GACAGCGCTTACCGCGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 784

QY 825 GTGCTGAGCGCTACCTTCCGAAGCTTTGATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 884

DB 785 GTGCTGAGCGCTACCTTCCCGAGCTTTTGAAGTGTGCTGCGTCTGCGAGGAGCGCGGAGCGAG 844

QY 885 CTGGTACCGTGTATGATAGCTTGGAGCCCGCATGGAACCCCAAGCTGTGGTGGTGGTGGTGGT 944

DB 845 CTGGTACCGTGTATGATAGCTTGGAGCCCGCATGGAACCCCAAGCTGTGGTGGTGGTGGTGGT 904

QY 945 GGCACTTCTTCACTCTCTTCACTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004

DB 905 GGCACTTCTTCACTCTCTTCACTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964

QY 1005 ACCTGTGATAACCAATATCTGACCGCGGAGTCTCTGGCTTTGAGGCGAGTCTTCTTCCAGCTG 1064

DB 965 ACCTGTGATAACCAATATCTGACCGCGGAGTCTCTGGCTTTGAGGCGAGTCTTCTTCCAGCTG 1024

QY 1065 CCCAGATGAGCGAGCTGTGGGGCTTTTGGAGTGACACCCCAAGGAGCATTTAGCGAGCCCC 1124

DB 1025 CCTAGGATGAGCGAGCTGTGGAGGCGGCTTACGTAAGCCCGAGGAGCATTTCAACAGGCCCC 1084

QY 1125 TACTATCCAGGCGCACTACCGCGCCCAACATCAACTGACATGGAATATCAAGGTGCCCCAAC 1184

DB 1085 TACTATCCAGGCGCACTACCGCGCCCAACATCAACTGACATGGAATATCAAGGTGCCCCAAC 1144

QY 1185 AACCGGAACGTGAAGGTGGCTTCAAACTCTTCTATCTGGTGGAGCCCAACGATACAGTGT 1244

DB 1145 AACCGGAACGTGAAGGTGGCTTCAAACTCTTCTATCTGGTGGAGCCCGGCTGGCTGGC 1204

QY 1245 GGCTCTGACCAAGAGTATGTGGAGATCAACGGGGAGAGTACTCGCGGTGAGAGTCC 1304

DB 1205 GGCACTTCTGCGGAGGAGTGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGTCC 1264

QY 1305 CAGTTTGTGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCCACTTCTGATCACTCG 1364

DB 1265 CAGTTTGTGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCTGATCACTCG 1324

QY 1365 TACACGGACACCGGGTTCTTAGCTGAGTACCTCTCTCTTACGACTCCACAGACCGGTGCCCA 1424

Db 1325 TACACGACACCGGCTTCTTAGCTGAATACCTCTCTACGACTCCAGTGCACCCATGCCCG 1384
QY 1425 GGGATGTTTCATGTCGAAGACTGACGCTGTCATCGAAGGAACTGCGCTCGACGGCTGG 1484
Db 1385 GGGAGTTTCACTGCGCCACCGGGCGGTGTATCCGAGAGAGCTGCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCCGGATTATAGTATGAGCGTTTACTGCCGATGCAATGTCACCCACCACTTC 1544
Db 1445 GCCGACTGCACCGACACAGCATGAGCTCAATGACAGTTGGACGCGCGCCACCACTTC 1504
QY 1545 ACGTGCAAAACACAGTTCTGCAAGCCCTCTCTCGGCTCTGTGACATGTCACAGACTGT 1604
Db 1505 ACGTGCAAGAACAGTTCTGCAAGCCCTCTCTCGGCTCTGTGACATGTCACAGACTGT 1564
QY 1605 GGGACGGAAGTACGAGGAGGGCTGACGCTGTCTCTGCTGGAGTTTCAAGTGTTCAT 1664
Db 1565 GGAGACAACAGCAGACGAGCGGGTGCAGTTGTCCGCGCCACGACCTTCAGTGTTCAT 1624
QY 1665 GGGAAAGTGTCTCTCAGAGCAGAGGAGGTAATGGGAAAGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTCTCAGAGCAGAGGAGGTAATGGGAAAGACAACTGTGGAGATGGGTCT 1684
QY 1725 GACGAGCTTCATGTGACAGGTGAATGTCTCTCTGCTGCAACCAATATACCTACCGCTGC 1784
Db 1685 GACGAGCTTCCTCCCAAGGTGAAGCTGTCTCTCTGCTGCAACCAATATACCTACCGCTGC 1744
QY 1785 CAAATGCGCTCTGCTGAGCAAGGCAACCTCTGAGTGTGATGGGAAAGAGCACTGTAGC 1844
Db 1745 CTCAATGGCTCTGCTGAGCAAGGCAACCTCTGAGTGTGATGGGAAAGAGCACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAACTGTGATGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAGGACTGCGACTGTGGGTGCGGTCTTACGAGACAGGCTCGT 1864
QY 1905 GTGGTTGGTGGCAAGTACGAGAGGAGGAGTGGCCCTGGCAGGTGAGCCTCCAGCC 1964
Db 1865 GTGGTTGGGCGCAGATGCGATGAGGCGAGTGGCCCTGGCAGGTGAGCCTCCAGCT 1924
QY 1965 CTGGGCGAGGCGCACTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGGTCTTGCA 2024
Db 1925 CTGGGCGAGGCGCACATCTCGGCTGTCTCTCTCATCTCTCTCCCACTGGCTGGTCTTGCC 1984
QY 2025 GCTCATGCTTTTCAAGTATGAGAAATTTCAAGTATCTGACATACAGATGTGAGCGCC 2084
Db 1985 GCACACTGTACATGATGATGACAGAGGATTCAGTACTCTAGACCCCAAGCAGTGGAC 2044
QY 2085 TTCTGGGTCTGTGGACACAGAGCAAGCGAGTGCCTCTGGGCTGACAGGAGCTGAAGCTC 2144
Db 2045 TTCTGGGTCTGACAGACAGAGCGCAGCGCCCTGGGCTGACAGGAGCGAGGCTC 2104
QY 2145 AAGCGTATCATACCCACCTTCTTCAATGATTTTCACTTTCGATATGACATCGCCTTG 2204
Db 2105 AAGCGTATCATCTCCACCCCTTCTTCAATGATTTTCACTTTCGATATGACATCGCCTTG 2164
QY 2205 CTGAGCTGAGAGTCTGGGAGTACAGCAGCGTCTGTCGCGCCCATCTGCTGCTGAT 2264
Db 2165 CTGAGCTGAGAGTCTGGGAGTACAGCAGCGTCTGTCGCGCCCATCTGCTGCTGAT 2224
QY 2265 GCTACCCATCTTCTTCTGTCGAGGAGGCTTGGGTCTGAGCTGGGCGGACACAAA 2324
Db 2225 GCCTCCCATGCTTCTTCTGTCGAGGAGGCTTGGGTCTGAGCTGGGCGGACACCCCAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGTTGAGATCCGTGTCTATCAACAGAC 2384
Db 2285 TATGAGGCACTGGCGGCTGATCTCTCAAAAGGTTGAGATCCGCGTCTATCAACAGAC 2344
QY 2385 ACCTGTGAGAGCTCATGCGCAGCAGATCACCCAGAGTATGATGTGGGTCTTCTC 2444
Db 2345 ACCTGTGAGAGCTCATGCGCAGCAGATCACCCAGAGTATGATGTGGGTCTTCTC 2404
QY 2445 AGTGGGGGTGTGAGTCTCTGCAAGGTTGACTCTGTGGCGCCCTTGTCAAGCGCGAGAA 2504

Db 2405 AGCGGGCGGTGAGTCTCTCCAGGCTGATTTCCGGGGACCCCTCTCCAGCTGGAGCG 2464
QY 2505 GATGGCGGATGTTCCAGGCTGCTGTGTGCTGAGCTGGGTGAAAGCTGCGCTCAGAGAAC 2564
Db 2465 GATGGCGGATGTTCCAGGCTGCTGTGTGCTGAGCTGGGTGAAAGCTGCGCTCAGAGAAC 2524
QY 2565 AAGCCAGGCTGTACACAAGGCTCCCTGTAGTTTCGGGACTGGATCAAGAGACACCTGGG 2624
Db 2525 AAGCCAGGCTGTACACAAGGCTCCCTGTAGTTTCGGGACTGGATCAAGAGACACCTGGG 2584
QY 2625 GTATAGCAGATGAGCAGACAGCCGACCAACACACCAAGGATGCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGGG---GCCACCCAAATGTGTACCTGCGGGGCCACCCATCGTCCACC 2641
QY 2685 CTGGGATACAGAGAGAGAAACACTGACCAATTTATGTGTGGCTCCCCCCCCCAACACA 2744
Db 2642 CCAAGTGTGACG---CTTGAGGCTGGAGACTGGACCGCTGACTGCACAGCGGCC---CCAGA 2699
QY 2745 ACCGAGCTGTGAACCTGCACTCTTTCAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGCTCCAAAT 2735

RESULT 6
AAD13155
ID AAD13155 standard; DNA; 3147 BP.
XX AAD13155;
XX AC
XX AC
DT 16-OCT-2001 (first entry)
XX Human membrane-type serine protease (MTSP) 1 protease domain DNA.
DE Human; transmembrane serine protease; membrane-type serine protease;
KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytotstatic;
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW matriptase; ds.
XX Homo sapiens.
XX Key
PH Location/Qualifiers
FT 1865..2590
FT CDS
FT /*tag= a
FT /product= "Human transmembrane serine protease
FT (MTSP) 1 protease domain"
FT /note= "CDS does not include start codon"
FT /partial
XX WC200157194-A2.
XX 09-AUG-2001.
XX 02-FEB-2001; 2001WO-US03471.
XX 03-FEB-2000; 2000US-0179982.
XX 18-FEB-2000; 2000US-0183542.
XX 22-JUN-2000; 2000US-0213124.
XX 26-JUL-2000; 2000US-0220970.
XX 08-SEP-2000; 2000US-0657986.
XX 22-SEP-2000; 2000US-0234840.
XX (CORV-) CORVAS INT INC.
XX Madison EL, Ong BO, Yeh J;
XX WPI; 2001-488877/53.
XX P-PSDB; AAE06936.
XX Novel single chain polypeptide comprising protease domain of type-II
PT membrane-type serine protease or its catalytically active portion
PT useful for treating and preventing cancer and tumor
XX Example 6; Page 225-227; 256pp; English.
PS

XX The invention relates to transmembrane serine proteases and their
CC corresponding nucleotides and the protease domain of a type-II
CC membrane-type serine protease (MTSP?). MTSP is useful for identifying
CC compounds that modulate or inhibit its proteolytic activity and for
CC formulating a medicament for treating neoplastic disease. MTSP and
CC its corresponding nucleotides are useful in preventing or treating
CC tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC is useful as a diagnostic marker for tumour development, growth and/or
CC progression and as immunogens to generate antibodies that specifically
CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC gene therapy. The present sequence is a DNA encoding protease domain of
CC human MTSP1 protein (also called matritase).

XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 22; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCCCAAAACCATGGGTAGCAATCGGGCGCGCAAGCCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTTCGGGGTACATGGGAGGAGTCGGGCCCGCAAGGGCGAGGGGGCCCGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTCAAACTCCCGGCTAGAGAACATGAATGGCTTGGAGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTCAAACTCCCGGCTAGAGAACATGAATGGCTTGGAGAG 124
QY 165 GGTGTGAGTTCCTGCTGCGAACAATGCGAAGAAAGTGAGAGAGCGGCCCGAGCGC 224
Db 125 GCGTGTGAGTTCCTGCGCAGTCAACAAGGTCAAGAAAGGTGGAAGAGCATGCGCGGCGC 184
QY 225 TGGTGTGTGTGGGAGTGTCTGTCAGTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGTGTGTGTGGGAGTGTCTGTCAGTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGTGTGGCACTTCCATATTCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGTGTGGCACTTTCAGTAGTACCGGAGCTGCTGCTCCAGAAAGTCTTCAATGGCTAC 304
QY 345 AGGATCAAAATGAGATCTTCTGGATGCGTATGAGAACTCCACTCCAGAGTTTATC 404
Db 305 AGGATCAAAATGAGAAATTTGTGGATGCTCTACAGAACTCCAACTCCACTGAGTTTGA 364
QY 405 AGCTGCGCAGCGAGGTGAAGGAGCGCTGAAGTCTGTACATGAAGTCCCTGTCTCTG 464
Db 365 AGCTGCGCAGCGAGGTGAAGGAGCGCTGAAGTCTGTACAGGAGTCCCATCTCTG 424
QY 465 GGTCTCTACCAAGAGTGGGTGTAACTGCCCTTCAAGTGAAGGAGTGTATCGCCCTAC 524
Db 425 GGCCTCTTACCACAGAGTGGGTGTGACGGCTTTCAGCGAGGCGAGCGTCAATCGCTAC 484
QY 525 TACTGTTCAGAGTTTCAGATCCCCCAGACCTTCGAGAGAGTGTATCGGCCATGGCT 584
Db 485 TACTGTTCAGAGTTTCAGATCCCCCAGACCTTCGAGAGGAGGCGGAGCGTCAATGGCC 544
QY 585 GTGAGCGAGTGTAAATTTGCCACCGGAGCAGCGGCACTGAAATCTCTGCTCAACA 644
Db 545 GAGAGCGGTAGTTCATGCTGCCCGCGGGCGGCTCCCTGAAAGTCTCTTGTGGTCAAC 604
QY 645 TCTGTGTGGCTTCCCATTTGACCCAGAAATGTGAGAGGACTCAGGACCAACAGCTGC 704
Db 605 TCAGTGTGGCTTTCCTCCACCGGACTCCAAAACAGTACAGAGGAGCCAGGACCAACAGCTGC 664
QY 705 AGTTTTCCTTCATGATCCCATGTTGTCAGAGTGTACAGCTTCACTACCTTGGTTCCCC 764
Db 665 AGCTTTGGCTTCGACGCGCGCGGTGTGAGCTGTATGCGCTTTCACCAAGCGCGGCTCCCT 724
QY 765 AACAGTCTCTACCGCGGCGATGCGCGTGTCCAGTGGGTCTCTGCGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCGCTCATGCGCGCTGCCAGTGGGCGCTTGGCGGGAGCGCGCACTCA 784

QY 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCCGAGCTTTGACCTTGGCTCTCGGACGAGCGGCGGAGGAC 844
QY 885 CTGGTCAACCGTGTATGATGAGCCTGAGCCCCCATGGAACCCACACGCTGTGTGCGGCTGTGT 944
Db 845 CTGGTCAACCGTGTATGATGAGCCTGAGCCCCCATGGAACCCACACGCTGTGTGCGGCTGTGT 904
QY 945 GGCACCTTCTCAACCTCTCTACACCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTCTCAACCT 964
QY 1005 ACCTGTATACCAATATCTGACCGCGCACATCTCTGGCTTTGAGGCGCACTTCTCTCTCTCTCTCT 1064
Db 965 ACCTGTATACCAATATCTGAGCGCGGCATCTCTGGCTTTGAGGCGCACTTCTCTCTCTCTCTCT 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGCTTTTGTGAGTGACACCAAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGCGCGCTTACGTAAAGCCCGAGGAGCATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCGCACTACCCCGCCCAACATCACTGACATGGAATATCAAGGTGCGCCAAAC 1184
Db 1085 TACTATCCAGGCGCACTACCCCGCCCAACATTTGACTGACATGGAATATCAAGGTGCGCCAAAC 1144
QY 1185 AACCGGAACCTGAGGTGCGCTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1244
Db 1145 AACCGCATGTGAGGTGAGCTTCAATTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1204
QY 1245 GGTCTCTGACCAAGGATGATGAGAGTCAACCGGGAGAAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTTCT 1264
QY 1305 CAGTTTGTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1364
Db 1265 CAGTTTGTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1324
QY 1365 TACACGACACCGGGTTCTTAGCTGAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
Db 1325 TACACGACACCGGGTTCTTAGCTGAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1384
QY 1425 GGGATTTCTATGTGCAAGACTGGAACGCTGATCGGAAGAACTGCGCTGCGGACGGGTGG 1484
Db 1385 GGGCAGTTTCACTGCTGCGCACCGGGCGGTGTATCCGGAAGAGCTGCGCTGTGATGGGTGG 1444
QY 1485 GCAGACTGCGCGGATTTATGATGATGAGCGTTACTTGCAGTGAATGCCACCCACAGCTTC 1544
Db 1445 GCGCATGTCACCGACACAGCGATGAGTCAACTGCAAGTTGCGACGCGCGCCACAGTTC 1504
QY 1545 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTCTCTGGGTCTGTGACAGTGTCAACAGCTGT 1604
Db 1505 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTCTCTGGGTCTGTGACAGTGTCAACAGCTGT 1564
QY 1605 GGGGACGGAGTTCAGGAGGAGGCTGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1664
Db 1565 GGGGACAAACAGCAGCAGGAGGCTGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1624
QY 1665 GGGGAGTGTCTCTCTCTCAGAGCCAGAGTGTAAATGGGAAGCAAACTGTGGAGATGGGTCT 1724
Db 1625 GGGGAGTGTCTCTCTCTGAAAAGCAGCAGTGTCAATGGGAAGCAAACTGTGGGAGCGGTTC 1684
QY 1725 GACGAGGCTTCTATGTGACAGCGTGAATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1784
Db 1685 GACGAGGCTTCTCTCTCTGCGGAGGAGTGAAGCTGCTCACTTGTACCAAAACACACTCTCTCTCTCT 1744
QY 1785 CAAAATGGGCTCTCTCTGAGCAGGAGGCGCACTCTGAGTGTGATGGGAGAGCGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTCTCTGAGCAAGGGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1804
QY 1845 GATGGCTTCTCTCTCTGAGCAAAATCTGATGAGGCTGCGATCTCTCTCTCTCTCTCTCTCTCTCTCT 1904
Db 1805 GACGGCTCAGTGAAGAGGAGTCTGCGACTGTGGGCTGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1864
QY 1905 GTGGTTGGTGGCAGCAATGCGGACGAGGGCGAGTGGCGCTCTGGCAGGTTGAGCTCTCCAGCC 1964

Db 425 GGGCCCTACCAAGAGTGTGGCTGTGACGGCTTACGAGGGCAGCGTCAATCGCTTAC 484
QY 525 TACTGGTTCAGAGTTTCAGCATCCCGCCACATCTGGCAGAGAGTTGTATCGCGCCATGGCT 584
Db 485 TACTGGTTCAGAGTTTCAGCATCCCGCAGCACCCTGTGTGAGGAGCGCAGCGGTCAATGSCC 544
QY 585 GTGGAGCGAGTTGTAAATGCAATGCCACCGCAGCAGCGGCACTGAAATCCTTCTGTGCTAACA 644
Db 545 GAGGAGCGGTAGTATGCTGCCCCCGCGCGCGCTCCCTGAACTCCTTGTGTGCTACC 604
QY 645 TCTGTGGTGGCTTCCCATTTGACCCCGAGATGCTGACAGGACTCAGGACAAACAGCTGC 704
Db 605 TCAGTGGTGGCTTCCCGCAGCACTCCAAAACAGTACAGAGGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTGTGCTGATGCCATGTGTGACAGTGCACACGCTTCACTACCCCTGGCTTCCCG 764
Db 665 AGCTTGTGCTGACCGCCCGGTGTGAGCTGATGCGTTCACACGCGCGCTTCCCT 724
QY 765 AACAGTCCCTACCGCGGCATGCGCGCTGCGAGTGGTCTGCGGGGGGACGCGCACTCT 824
Db 725 GACAGCCCTACCGCGCTCATGCGCGCTGCCAGTGGGCGCTGCGGGGGGACGCGCACTCA 784
QY 825 GTGCTGAGCCTCACCTTCGAGAGCTTTGATGTGCTCCTGTGATGACCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCGAGAGCTTTGACCTTGCCTTCTGCGACGAGCGCGGAGCGAC 844
QY 885 CTGGTCACTGTATGATGAGCTGAGCCCGCATGGAACCCACGCTGTGTGCGGCTGTGT 944
Db 845 CTGGTGAAGGTGTACACACCTGAGCCCGATGAGCCCGGAGCGCGCTGTGAGTGTGT 904
QY 945 GGCACCTTCTCACCTCTCAAACTGACTTTCCTCTCCCGAGAAAGCTTCTCTTGTGTC 1004
Db 905 GGCACCTTACCTCTCTCAAACTGACTTTCCTCTCCCGAGAAAGCTTCTCTCTCATC 964
QY 1005 ACGCTGATAACCAATACGACCGCGGACATCGCTGGCTTTGAGGCGCACTTCTTCAGCTG 1064
Db 965 ACATGATTAACCAACATGAGCGCGGCGATCCCGGCTTTGAGGCGCACTTCTTCAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCCAAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCGCTTACGTAAAGCCAGGAGGACATTCACACGCCCC 1084
QY 1125 TACTATCCAGCCACTACCGCCCAACATCACTGCAATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGCCACTACCCACCAACATGACTGCAATGGAATATGAGGTGCCCAAC 1144
QY 1185 AACCGGAACGTGAAGTGTGCGCTTCAAACTCTTCTATCTGTGTGGACCCCAAGTACAGTG 1244
Db 1145 AACCATGATGAAGGTGAGCTTCAAACTCTTCTATCTGTGTGGACCCCAAGTACAGTG 1204
QY 1245 GGCTCTCTGACCAAGGACTATGTGGAGATCAACCGGGAGAAAGTACTCGGTGAGAGGTCC 1304
Db 1205 GGCACCTTCCCGCAAGGACTAGTGTGAGATCAATGGGAGAAATACTCGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGACAGCAACAGCAGCAAGATACAGTCCACTTCCATTTGATCACTCG 1364
Db 1265 CAGTTCTGTGTGACAGCAACAGCAGCAAGATACAGTTCGCTTCCACTCAGATCAGTCC 1324
QY 1365 TACACGACACCGGGTTCCTAGCTGAGTACCTCTCTACGACTCCACAGCCCGTCCCA 1424
Db 1325 TACACGACACCGGGTTCCTAGCTGAGTACCTCTCTACGACTCCAGTACCCCATGCCCCG 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGGAAGTGTGCAATCCGAAAGGAACTGCGGTGCGAGCGGTGG 1484
Db 1385 GGGCAGTTTCAGTCCGCAACCGGGCGGTGTATCCGGAAGGAGCTGCGGTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCGGATTTAGTGTGAGGCTTACTGCGGATGCAATGCCACCCACAGTTC 1544
Db 1445 GCCAGCTGCACCGACCAAGCGATGAGCTCAACTGCGAGTTGCGACGCGCGGCCACAGTTC 1504
QY 1545 ACGTGCAAAACCAAGTTCGCAAGCCCTCTCTGCGGTCTGTGACAGTGTCAACAGCTGT 1604

Db 1505 ACGTGCAAAACCAAGTTCGCAAGCCCTCTCTCTGGGTCTGCGACAGTGTGAACGACGTGC 1564
QY 1605 GGGACACGAAGTACGAGGAGGCTGAGCTGCTCTGCTGGAGTTTCAAGTGTTCCTCAAT 1664
Db 1565 GGAACAAACAGCAGCAGAGCGGGGTGAGTTGTCCGGCCAGACCTTCAGGTGTTCCTCAAT 1624
QY 1565 GGAAGTGTCTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGAAGTGTCTCTCGAAAGCCAGCAGTGCATGGGAAGGACGACTGTGGGAACGGGTCC 1684
QY 1725 GACAGAGCTTCATGTGACAGCGTGAATGTGCTCTTTGACACCAATATATACCTACCGTGC 1784
Db 1685 GACAGAGCTCTCTCCCGCAAGGTGAACGTGTCTCACTTTGTACCAACACACCTACCGTGC 1744
QY 1785 CAAAATGGCTCTCTCTGAGCAAGGGCAACCTGAGTGTGATGGGAAGACGAGCTGTAGC 1844
Db 1745 CTCAATGGCTCTCTGTGAGCAAGGGCAACCTGAGTGTGACGGGAAGGAGCTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAATACTGTGACTGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAAGGACTGCGACTGTGGGTCTGCGGTCAATTCACGAGACAGGCTCGT 1864
QY 1905 GTGGTGTGCGCAGCAATGCGGACGAGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1865 GTTGTGGGGGACGAGTGGCGATGAGGCGAGTGGCCCTGGCAGGTAAAGCTCGATGCT 1924
QY 1965 CTGGGCGAGGGCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGCTGCTCTCTGCA 2024
Db 1925 CTGGGCGAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCAACTGCTGGTCTCTGCC 1984
QY 2025 GCTCATTTGCTTTTCAAGATGACAAATAATTTCAAGTACTCAGACTACAGATGTGGACGCC 2084
Db 1985 GCACACTGTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAACGAGTGGACGCC 2044
QY 2085 TTCTGCGGTCTGTGAGCAGCAGCAAGCGCAGTGCCTCTGGGGTGGCAGGAGCTCAAGCTC 2144
Db 2045 TTCTGCGGTGTGACGACCCAGAGCCAGCGCGCCCTCTGGGGTGGCAGGAGCTC 2104
QY 2145 AAACGTATCATCACCCCTTCTCTCAATGATTTTCACTTTCGACTATGACTCGCTTG 2204
Db 2105 AAGGCACTCATCTCCACCCCTTCTTCAATGACTTTCACCTTGGACTATGACTCGCGCTG 2164
QY 2205 CTGAGCTGAGAGTGTGGAGTACAGCAACCGTCTGCGCCCTCATCTGCTGCTGCTGAT 2264
Db 2165 CTGAGCTGAGAGAAACCGCAGAGTACAGCTCCATGTGTGGGCCCATCTGCTGCGCGAC 2224
QY 2265 GCTACCCATCTTCTCTGCTGCGCAAGGCCATCTGGGTGACAGCTGCGGGGACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCTCTGCGGCAAGGCCATCTGGGTGACAGGCTGCGGGACACACCCAG 2284
QY 2325 GAGGGAGTACCGGAGCGCTGATCTGCAAGAGGTGAGATCCGTGTGCTCATCAACGAGACC 2384
Db 2285 TATGGAGCACTGGCGCTGATCTCTGCAAAAGGTGAGATCCGCTCATCAACGAGCC 2344
QY 2385 ACCTGTGAGACCTCATGCGCAGAGATCAACCCACGAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGCGAGAACCTCTCTGCGCAGCAGATCAACGCGCATGATGTGCTGCGGTCTCCTC 2404
QY 2445 AGTGGGGTGTGACTCTGCGCAGGCTGACTGTGTGGCCCTTGTCAAGCGGGAGAAA 2504
Db 2405 AGCGCGCGGTGGAACCTCTCCAGGGTGTGTCGGGGGACCCCTGTCCACGCTGGAGGCG 2464
QY 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGGCGATCTTCCAGGCGCGGTGTGTGAGCTGGGAGACGCGCTCAGAGGAAC 2524
QY 2565 AAGCAGGCGGTACACAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCAGACTGGG 2624
Db 2525 AAGCAGGCGGTACACAGGCTCCCTCTGTTTCGGGACTGGATCAAGAGAACTCTGGG 2584
QY 2625 GTATGACGATTCGACAGCAGCGGACCAACACCCACGAGGATGCCCGCATGACCA 2684
Db 2585 GTATAGGGCGGGG---GCCACCAAAATGTGTACACCTGCGGGGCGCCACCATCGTCCACC 2641

Db 1943 GGCACCTGCCCCAAGGACTAGTGGAGATCAATGGGAGAAATACTGCGAGAGAGTCC 1884
QY 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATCTTGATCACTCG 1364
Db 1883 CAGTTGTGTCACAGCAACAGCAGCAAGATTACAGTCCACTTCCATCTTGATCACTCG 1824
QY 1365 TACACGACACCGGTTCTTAGCTGAGTACCTCTCTACGACTCCAAAGACCCGTGCCCA 1424
Db 1823 TACACGACACCGGTTCTTAGCTGAGTACCTCTCTACGACTCCAAAGACCCGTGCC 1764
QY 1425 GGGATGTTCACTGCAAGACTGAGCGGTGCATCCGAAGGAACCTGCGCTGCGACGCGTGG 1484
Db 1763 GGGCAGTTTCAGTGCAGCGCAGCGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1704
QY 1485 GCAGACTGCCCGGATTATAGTGTATGAGCGTTTACTGCGGATGCAATGCCACCCACAGTTTC 1544
Db 1703 GCCGACTGCACCGACCAAGCAGTGCAGTCACTGAGTGGAGCGCGGCCACAGTTTC 1644
QY 1545 ACGTGCAAAAACCGATTCTGCAAGCCCTCTTCTGGGTCTGTGAGTGTCAAGACTGT 1604
Db 1643 ACGTGCAAAAACCGATTCTGCAAGCCCTCTTCTGGGTCTGTGAGTGTCAAGACTGT 1584
QY 1605 GGGGACGGAAGTACGAGGAGGCTGCACTGTCTGCTGGAGTTTCAAGTGTTCGAT 1664
Db 1583 GGAGACAACAGCGACGAGCGGGGTGCAAGTGTCTCGGCCACAGCTTCCAGTGTTCGAT 1524
QY 1665 GGGAGTGTCTTCCCTCAGAGCCAGAGTGTAAATGGGAAGGACAACCTGTGAGATGGGTCT 1724
Db 1523 GGGAGTGTCTTCCCTCAGAGCCAGAGTGTAAATGGGAAGGACAACCTGTGAGATGG 1464
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTGCTCTTGTGCAACCAATATACCTACCGTGC 1784
Db 1463 GACGAGGCTTCCTGCCCCAAGTGAAGTGTGCTCTTGTACCAAAACACACCTACCGTGC 1404
QY 1785 CAATAAGTCTCTCTGAGCAAGGCAACCTGAGTGTGATGGAGACGAGTGTAGC 1844
Db 1403 CTAAATGGGCTCTCTGAGCAAGGCAACCTGAGTGTGATGGAGAGGAGTGTAGC 1344
QY 1845 GATGGCTCCGATGAGAAAACCTGTGAGTGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
Db 1343 GACGGCTCAGATGAGAGGACTGCGACTGTGGGTGCGGTCAATTCAGAGACAGGCTCGT 1284
QY 1905 GTGTTGGTGGCAGCAATGCGGACGAGGCGAGTGGCCCTGCGAGGTGAGCTCCAGGCC 1964
Db 1283 GTTCTGGGGCAGCGATGCGGATGAGGGGAGTGGCCCTGGCAGGTAAAGCCTCATGCT 1224
QY 1965 CTGGGCCAGGGCCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGTTCTGCA 2024
Db 1223 CTGGGCCAGGGCCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGTTCTGCA 1164
QY 2025 GCTCAATGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1163 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAACGAGTGGACGGCC 1104
QY 2085 TTCTCTGGGTCTGCTGAGACGAGCAAGCGGAGTGGCTCTGGGGTGGCAGGAGTGAAGCTC 2144
Db 1103 TTCTCTGGGTCTGACGACACGAGCAGCGAGCGCCCTGGGGTGGCAGGAGCGAGGCTC 1044
QY 2145 AAAGTATCATCAACCCCTTCTTCAATGATTTCACTTCCGACTATGATGATGCGCTTG 2204
Db 1043 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTCCGACTATGATGATGCGCTG 984
QY 2205 CTGAGCTGGAGAACTCGGTGAGTACAGCACCCCTCGTGGGCCCCCTCTGCTGCTGCTGAT 2264
Db 983 CTGAGCTGGAGAAACCGGACAGATAAGCTCCATGATGGGGCCCTCTGCTGCTGCTGAT 924
QY 2265 GCTACCATGTTCTTCCCTGTGCAAGGCGATCTGGGTGCAAGGCTGGGGGCAACAAAA 2324
Db 923 GCTTCCCATGTTCTTCCCTGTGCGGCAAGGCCATCTGGGTGCAAGGCTGGGGGCAACCCAG 864
QY 2325 GAGGGAGTACCGAGCGCTGATCTCAGAGAGGTTGAGATCCGTGTCATCAACAGGAC 2384

Db 863 TATGAGGCACTGGCGCGCTGATCTGCAAAAGGCTGAGATCCCGCTCATCAACAGGACC 804
QY 2385 ACCTGTGAGGACCTCATGCGCGACAGATCAACCCCAAGATGATGTGTGGTTCCTC 2444
Db 803 ACCTGCGAGAACTCTCTGCGCGACAGATCAACCGCGCATGATGTGTGGTTCCTC 744
QY 2445 AGTGGGGGTGTGGACTCTCTGCGAGGCTGACTCTGCTGGTGGCCCTTGTCAAGCGCGAGAAA 2504
Db 743 AGCGCGCGGTGGACTCTCTGCGAGGCTGATTCGCGGGGACCCCTGTCCAGGTGGAGCG 684
QY 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 683 GATGGGCGGATCTTCCAGGCGGTGTGTGTGAGCTGGGGAGACGCTGCGCTCAGAGGAAC 624
QY 2565 AAGCCAGCGGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGAACTGCTGG 2624
Db 623 AAGCCAGCGGTGTACACAAGGCTCCCTGTGTTCGGGACTGGATCAAGAGAACTGCTGG 564
QY 2625 GTATAGCAGCATGTGACAGACAGCGCGACCAAAACACACAGGAGTCCCGACATGCACA 2684
Db 563 GTATAGGCGCGGG---GCCACCAAAATGTGTACCTGCGGGGCCACCCATCGTCCACC 507
QY 2685 CTTGGATACAGGAGAGGAACAACCTGACGACATTTATGTGTGGCTCCCGCCCCCAACACA 2744
Db 506 CCACTGTGTGACG-CCTGAGGCTGGAGACTGGACCGCTGACTGACCCAGCGGCC-CCAGA 449
QY 2745 ACCAGACTGTGAGTGCATCCTTAGGACTCAGAGT 2780
Db 448 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 413

RESULT 9

AAL53444

ID AAL53444 standard; DNA; 3147 BP.

XX AAL53444;

DT 12-DEC-2002 (first entry)

XX Type II transmembrane serine protease 1 coding DNA SEQ ID No 1.

XX Cytostatic; type-II membrane-type serine protease 7; MtsP7; malignancy;
XX neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
XX malignant; enzyme; gene; ds.

XX Homo sapiens.

XX WO200272786-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US07903.

XX 13-MAR-2001; 2001US-275592P.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO;

XX WPI; 2002-732827/79.

XX P-PSDB; AAO22929.

XX New transmembrane serine protease 7 (MtsP7) polypeptide for diagnosing
XX neoplastic diseases, monitoring tumor progress or therapeutic
XX effectiveness, or identifying MtsP7 modulators for treating tumors or
XX cancers

XX Disclosure; Page 168-172; 184pp; English.

XX The invention relates to a purified single or two-chain polypeptide,
XX which comprises the protease domain of a type-II membrane-type serine
XX protease 7 (MtsP7) or its catalytically active portion. The polypeptide
XX comprising MtsP7 is useful for detecting or diagnosing a neoplastic

QY 1965 CTGGCCAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGCTGGCTGTTCTGCA 2024
Db 1925 CTGGCCAGGGCCACATCTCGGTGCTTCCCTCATCTCTCCCACTGGCTGGTCTCTGCC 1984
QY 2025 GCTCAATGCTTTTTCAGGATGACAAATTTTCAAGTACTCAGACTACAGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATGATGACAGAGGATTTAGGTACTCAGACCCACGCGAGTGGACGGCC 2044
QY 2085 TTCTGTGGTCTGTGGACCAAGAGCAAGCGCAGTGTCTCTGGGTGTCAGGAGCTGAAGCTC 2144
Db 2045 TTCTGTGGTCTGTGGACCAAGAGCAAGCGCAGTGTCTCTGGGTGTCAGGAGCTGAAGCTC 2104
QY 2145 AAGAGTATCATACCCACCTCTCTTCAATGATTTACCTTTCAGTATGATGATGCTCTTG 2204
Db 2105 AAGCGCATCATCTCCACCTCTCTTCAATGATTTACCTTTCAGTATGATGATGCTCTTG 2164
QY 2205 CTGGAGCTGGAGAGTCTGGTGGAGTACAGACACGCTGTGTGCGCCCACTCTCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACGGGAGAGTACAGTCTCATGTGTGCGCCCATCTCTGCTGCGGAC 2224
QY 2265 GCTACCATGCTTCTCTGTGTGGCAAGGCCATCTGGGTACAGGCTGGGGGACACAAAA 2324
Db 2225 GCTCTCCATGCTTCTCTGTGCGGCAAGGCCATCTGGGTACAGGCTGGGGGACACCCAG 2284
QY 2325 GAGCGAGGTACCGAGGCTGATCTCTGCAAGAGGTGAGTCCGTGTCTATCAACCCAGACC 2384
Db 2285 TATGGAGGCTCTGGCGGCTGATCTCTGCAAGAGGTGAGTCCGTGTCTATCAACCCAGACC 2344
QY 2385 ACCTGTGAGGACCTCTATCGCGCAGCAGATCACCCACGATGATGTGTGGGTCTTCTC 2444
Db 2345 ACCTGTGAGAACCTCTCTGCGCAGCAGATCACCGCGCATGATGTGTGGGTCTTCTC 2404
QY 2445 AGTGGGGTGTGACTCTCTGCCAGGTGACTCTGTGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGGGGTGTGACTCTCTGCCAGGTGATTCGGGGGACCCCTGTCCAGCTGGAGGG 2464
QY 2505 GATGGCGAATTTTCAGGCTGTGTGTGAGCTGGGTGAAGCTGCGCTCAGAGAAC 2564
Db 2465 GATGGCGGATCTTCAGGCGGTGTGTGTGAGCTGGGGAGCGGCTGCGCTCAGAGAAC 2524
QY 2565 AAGCCAGGCTGTACAAAGGCTCTCTGTAGTTTCGGGACTGGATCAAGAGCACACTGGG 2624
Db 2525 AAGCCAGGCTGTACAAAGGCTCTCTGTTCGGGACTGGATCAAGAGCACACTGGG 2584
QY 2625 GTATAGCAGATGGACAGACGCGGACCAACACCCACAGGATGCCCGACATCCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCAATGTGTACACCTGCGGGGCCACCATCGTCCACC 2641
QY 2685 CCTGGATACAGGAGAGAACTGACGACATTTATGCTGTGGCTCCCGCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGCGAGGTGGAGCTGGACCGCTGACTGACACAGCGGCC-CCAGA 2699
QY 2745 ACCAGACTGTGAATGATCTCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 10

AAL53445

ID AAL53445 standard; DNA; 3147 BP.

XX AAL53445;

XX AAL53445;

DT 12-DEC-2002 (first entry)

XX Type II transmembrane serine protease 1 domain DNA SEQ ID No 3.

XX DE type-II membrane-type serine protease 7; MTSP7; malignancy;

XX KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;

XX KW malignant; enzyme; gene; id.

XX OS Homo sapiens.

XX

PN WO200272786-A2.
XX 19-SEP-2002.
XX 13-MAR-2002; 2002WO-US07903.
XX 13-MAR-2001; 2001US-275592P.
XX (CORV-) CORVAS INT INC.
PA Madison EL, Ong EO;
PI WPI; 2002-732827/79.
DR P-PSDB; AAO22930.
XX New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
neoplastic diseases, monitoring tumor progress or therapeutic
effectiveness, or identifying MTSP7 modulators for treating tumors or
cancers
XX Disclosure; Page 174-176; 184pp; English.
XX The invention relates to a purified single or two-chain polypeptide,
which comprises the protease domain of a type-II membrane-type serine
protease 7 (MTSP7) or its catalytically active portion. The polypeptide
comprising MTSP7 is useful for detecting or diagnosing a neoplastic
disease, a pre-malignant lesion, a malignancy or other pathologic
condition in a subject. This polypeptide is also useful for monitoring
tumour (e.g. tumour of the breast, cervix, prostate, lung, ovary or
colon) progress and/or therapeutic effectiveness. The inhibitor of the
polypeptide or the inhibitor of the polypeptide's zymogen form is useful
for treating or preventing a neoplastic disease, or tumour initiation.
CC growth or progression, or a (pre-)malignant condition. The polypeptide or
polynucleotide is also useful for identifying modulators of MTSP7, which
may be used to treat cancers or tumours. This polynucleotide sequence
represents the coding DNA for a protein domain of the type-II membrane-
type serine protease 1 relating to the invention.
XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match

Best Local Similarity 60.6%; Score 1883.2; DB 24; Length 3147;

Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAAACCATGGTGTAGCAATCGGGCCGCAAGCGCGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCGCAAGGCGAGGGGCCCGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCAAGGAGAGTGAATGGCTTTGAGGAG 124
QY 165 GCTGTGGAGTTCCTGCTCGGACATGCCAAGATGCCAAGATGGAGAGGCGGCCCGAGCG 224
Db 125 GCGGTGGAGTTCCTGCCAGTCAACAACTGCAAGAGGTGGAAAGCATGGCCCGCGGCGC 184
QY 225 TGGGTGTGTCTGGTGGCAGTGTCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGTGTCTGGTGGCAGTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGTGTGGGACTTCCATTTATCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGTGTGGGACTTTCAGTACCGGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 304
QY 345 AGGATCACAATAGATCTTTCTGATGCGGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCACAATAGAAATTTTGTGATGCGCTACGAGAACTCCCACTCCACTGAGTTGTA 364
QY 405 AGCTGGCCAGCAGGTGAAGAGGCGCTGAAGCTGTGTAAGTGAAGTCCCTCTCTCTG 464
Db 365 AGCTGGCCAGCAGGTGAAGAGCCTGAGCTGTGTAAGTGTGTGTGTGTGTGTGTGTGTGT 424
QY 465 GGTCTCTACCAAGAGTGGCTGTAACTGCTTCTAGTGGGGCAGTGTGTCTATCGCTTAC 524

Db 425 GGGCCCTACCAAGAGTGGCTGTGACGGCTTTCAGCGGGCAGCGTCATCGCCTAC 484
QY 525 TACTGGTCAGAGTTACAGATCCCCCACAACCTGGCAGAGAGTTGATCGGGCCATGGCT 584
Db 485 TACTGGTCAGATTCAGATCCCGCAGCACTGGTGAGGAGCGCGCGCTCATGGCC 544
QY 585 GTGAGGAGTTGTAACTTCCACCCCGAGCAGCGGCACTGAAATCTTGTGCTAAACA 644
Db 545 GAGGAGCGGTAGTATCATCTGCCCGCGCGCGCTCCCTGAAAGTCTCTTTGTGTCAAC 604
QY 645 TCTGTGTGGCTTCCCATTTGACCCAGATGTGACAGGACTCAGAGCAACAGCTGC 704
Db 605 TCAGTGTGGTTTTCCTCCACGGAATCCAAACAGATCAGAGACCCAGGCAACAGCTGC 664
QY 705 AGTTTTCCTGATGCCATGTGTGACAGCTGACAGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCTGACGCGCGCGGTGTGAGCTGATGGCTTCCACACGCGGCTTCCCT 724
QY 765 AACAGTCTCTACCCGGCGCATGCCCGTGCAGTGGGTCTGTGCGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGTCCCGCTGCGCAGTGGCGCTTGGCGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCTTCACTTCCGAAGCTTTGATGTGCTTCCCTGCTGATGAGCATGCGAGTGAC 884
Db 785 GTGCTGAGCTTCACTTCCGAGCTTTTGAACCTTGGCTTCCGACGAGCGCGCAGCGAC 844
QY 885 CTGCTCACCGTGTATGATAGCTGAGCCCATGAGAACCCACGCTGTGTGGCGCTGTGT 944
Db 845 CTGCTGAGCGGTGTAAACAACCTTGAAGCCCATGAGCGCCCATGCGCTGTGTGTGT 904
QY 945 GGCACCTTCTCACCTCTTACACCTGACTGATCTTCTCTTCCAGAAAGCTTCTTGTGTC 1004
Db 905 GGCACCTTACCTTCCCTCTTCAACCTGACCTTCCACTCTTCCAGAAAGCTTCTGCTCATC 964
QY 1005 ACGTGATAACCAATACTGACCGCGACATCTCTGGCTTTTGAAGTGACACCCAAAGGACATTTAGCAGCCCC 1064
Db 965 AACTGATAACCAACTGAGCGCGCATCTCCGGCTTTTGAAGTGACATTTTCCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGCTTTTGAAGTGACACCCAAAGGACATTTAGCAGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGCGCGTTACGTAAAGCCACCGGGGACATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGCCCACTACCGCCCAACATCAATGACATGAGCAATGAGGTTCAAGGTGCCAAC 1184
Db 1085 TACTACCCAGGCCCACTACCCCAACATTTGACTGCACATGGAACATTTAGGTGCCAAC 1144
QY 1185 AACCGAAAGTGAAGGTGCGCTTCAAACTCTTCTATCTGTGTGACCCCAAGTACCACTG 1244
Db 1145 AACGAGATGTGAAGTGAAGTTCAAATTTCTTACCTGTGAGCGCGCGCTGCGTGG 1204
QY 1245 GGCTCTTGCAACAGGACTATGTGGAGATCAACGGGAGAAAGTACTGCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGAGAAATACTTGGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAGCAGCAAGATTTACAGTCCACTTCCATTTGATCACTG 1364
Db 1265 CAGTTTGTGTGACAGCAGCAGCAAGATTTACAGTTCGCTTCCACTCAGATCAGTCC 1324
QY 1365 TACACGACACCGGGTTCCTAGTGTGAGTACTCTCTTACGACTCCCAAGCAGCCGTCGCCA 1424
Db 1325 TACACGACACCGGGTTCCTAGTGTGATCTCTTACGACTCCCAAGTACCCATGCCCG 1384
QY 1425 GGGATTTGATGTGCAAGACTGAGCAGGTGATCCGAAAGGAACTGCGCTGCGAGCGGTG 1484
Db 1385 GGGCAGTTTCACTGTCGCCAGCGGGCGGTGTATCCGGAAGGAGCTGCGTGTGTGCTG 1444
QY 1485 GCAGACTGCCGATTTATGATGATGAGCTTACTGCGATGCAATGCCACCCAGCTTC 1544
Db 1445 GCCGACTGCAACCCAGCAGCAGTGAAGTCAACTGAGTTGCGACCGCGGCCACCACTTC 1504
QY 1545 ACGTGCAAAACCAAGTCTTCAAGCCCTTCTTGGGTCTGTGACAGTGTCAACGACTGT 1604

Db 1505 ACGTGCAAAACCAAGTCTTCAAGCCCTTCTTGGGTCTGCGCAGCTGTGTAACGACTGC 1564
QY 1605 GGGGACGGAAGTGCAGGAGGGCTGCAGCTGTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGAACAACAGCAGCAGGAGGGGTGCAGTTGTTCGGGCCAGACCTTTCAGGTGTTCGAAT 1624
QY 1665 GGGAAAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTTCGAAAAGCCAGCAGTGCATTTGGGAGAGCAGCTGTGGGAGCGGTTC 1684
QY 1725 GACGAGCTTCATGTGACAGCGTGAATGTCTTCTTGGCAACAAATATACTACCGCTGC 1784
Db 1685 GACGAGCGCTTCTGCCCGCAGGTGAACCTGCTCCTTGTACCAACACACACTTACCGCTGC 1744
QY 1785 CAAAATGGCGCTCTGTGTGACAAAGGCAACCTCAGTGTGTGATGGGAGAGCGGACTGTAGC 1844
Db 1745 CTCAATGGGTCTGTGTGACAAAGGCAACCTTGTGATGTGACGGGAGGAGGACTGTAGC 1804
QY 1845 GATGGCTTCGATGAGAAAACTGTGACTGTGGCTGCGATCTTTCACAAACAGGCTTCGC 1904
Db 1805 GACGGCTCAGATGAGAAAGGACTTGCAGCTGTGGCTGGGTCTTTCAGCAGACAGGCTCGT 1864
QY 1905 GTGGTTGGTGGCAGCAATGCGGACGAGGGGAGTGGCCCTTGGCAGAGTGAGCCCTCAGGCC 1964
Db 1865 GTTGTGGGGGCGACGATGCGGATGAGGGCGAGTGGCCCTTGGCAGGTAAAGCTGTGCT 1924
QY 1965 CTGGGCCAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGTCTGTGCA 2024
Db 1925 CTGGGCCAGGGCCCACTTGTGGGTCTTCCCTCATCTCTCCCAACTGGCTGTGCTCTGCC 1984
QY 2025 GCTCATTTGCTTTTCAAGTACTCAGACTTACACGATGTGAGAGCGGCC 2084
Db 1985 GCACACTGTCTACATGATGACAGGATTCAGGTACTCAGACCCACGCGATGAGCGGCC 2044
QY 2085 TTTCTGGGTCTGTGACACGAGCAGCGAGTGGCTCTGGGTGAGGAGCTGAAGCTC 2144
Db 2045 TTTCTGGGTCTGTGACACGAGCAGCGCCCTTGGGTGAGGAGCGAGGCTC 2104
QY 2145 AAAGTATCATACCCACCTTCTTCAATGATTTCACTTGCAGCTTACAGATGTGAGAGCGGCC 2204
Db 2105 AAGGGATCATCTCCACCCCTTCTTCAATGATTTCACTTGCAGTATGACATCGCGCTG 2164
QY 2205 CTGGAGCTGGAGAAAGTTCGGTGGAGTACAGCAGCTGTGGCGCCCACTTGCCTTCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGCTCCATGTGTGGCGCCCATCTGCTTCCCGGAC 2224
QY 2265 GCTACCCATGTCTTCTCTGCGGAGGCACTTGGGTACAGCTGGGGGACACAAA 2324
Db 2225 GCTCCCATGTCTTCTCTCGGCAAGGCACTTGGGTACAGGCTGGGGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTTGCAGAAAGGTGAGATCCGTCTCATCAACAGACC 2384
Db 2285 TATGGAGCACTGGCGCTGATCTTGCAGAAAGGTGAGATCCGCTCATCAACAGACC 2344
QY 2385 ACGTGTGAGCACTCATGCGGACGAGATCACCCAGAAATGATGTGTGGTTCCTC 2444
Db 2345 ACGTGTGAGCACTCATGCGGACGAGATCACCGCGCATGATGTGTGGTTCCTC 2404
QY 2445 AGTGGGGTGTGCACTCTTCCAGGCTGACTGTGTGGCCCTTGTCAAGCGGGAGAA 2504
Db 2405 AGCGGGCGGTGCACTCTTCCAGGCTGATTCGGGGGACCCCTTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGCAATGTTCAGGCTGTGTGAGTGTGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATCTTCCAGCGCGTGTGTGAGTGTGGGAGAGCGGTGCTCAGAGGAAC 2524
QY 2565 AAGCAGGCGGTGTACAAAGGCTTCTGTGTGAGTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 2525 AAGCAGGCGGTGTACAAAGGCTTCTGTGTGAGTTCGGGACTGGATCAAGAGAACTGGG 2584
QY 2625 GTATAGCAGTGTGACAGCAGCGCCAGCAACACCCAGGGATGCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTTGGCGGGGCAACCCATCGTCCACC 2641

QY 2685 CCTGGATCAGGAGGAGAACATGACGACATTTATGCTGTGGCTCCCGCCCAACACA 2744
 DB 2642 CCAGTGTGCACG-CCTGCGAGGTGGAGACTGGACCGCTGACTGACACAGCGCC-CCAGA 2699
 QY 2745 ACCGACTGTGAAGTGCATCCTTAGGACTCAGAGT 2780
 DB 2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCMAAT 2735

RESULT 11

ABZ58500

ID ABZ58500 standard; cDNA; 3147 BP.

XX AC ABZ58500;

XX DT 13-MAY-2003 (first entry)

XX DE Transmembrane serine protease 1 (MTSP1) cDNA.

XX KW Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;
 XX KW Cytostatic; dermatological; cardiant; vulnerary; ophthalmological;
 XX KW gene therapy; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 23..2590

FT FT /*tag= a

XX FT /product= "MTSP1"

XX PN WO2003004681-A2.

XX PD 16-JAN-2003.

XX PF 03-JUL-2002; 2002WO-US21208.

XX PR 03-JUL-2001; 2001US-302939P.

XX PA (CORV-) CORVAS INT INC.

XX PI Madison EL, Ong EO;

XX DR WPI; 2003-239207/23.

XX PS P-PSDB; ABP72376.

XX PT New type-II membrane-type serine protease 20 polypeptides, useful for
 XX PT preparing a medicament for diagnosing, treating or preventing cancer,
 XX PT dermatological disorders, aberrant wound repairs or crest syndromes -
 XX PS Disclosure; Page 194-198; 216pp; English.

CC The present sequence is that of cDNA encoding human type II
 CC transmembrane serine protease 1 (MTSP1). The invention relates
 CC to novel human type II transmembrane serine protein 20 (MTSP20)
 CC (see ABP72374) and nucleic acids encoding it (see ABZ58499).
 CC Also claimed are methods of inhibiting tumour initiation, growth
 CC or progression by inhibiting MTSP20 activity, and of treating or
 CC preventing a disease or disorder associated with undesired and/or
 CC uncontrolled angiogenesis or neovascularisation, especially
 CC malformations and cardiovascular disorders, chronic inflammatory
 CC diseases, aberrant wound repairs, circulatory disorders, crest
 CC syndromes, dermatological disorders and ocular disorders using
 CC an inhibitor of MTSP20.

SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
 Best Local Similarity 81.2%; Pred. No. 0;
 Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCATTGGGCGCGCAGGCCGGAGGGGCTCTCAG 104

DB 5 GAGCGCCCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCCCCGAAG 64
 QY 105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGCTAGAGAAATGAATGGCTTTGAGGAG 164
 DB 65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGCTAGAGAAATGAATGGCTTTGAGGAG 124
 QY 165 GGTGTGGAGTTCCTGCTCGGAACAATGCCAAGAAAGTGGAGAAGCGAGCGCCACAGCGC 224
 DB 125 GCGGTGGAGTTCCTGCCAGTCAACAACGTCAAGAGGTGGAAGAGCATGGCCCGGGGCGC 184
 QY 225 TGGGTGGTGTGGTGGAGTGTCTTTCAGCTTCCTTGTCTCTCCCTCAATGCTGGCTTTC 284
 DB 185 TGGGTGGTGTGGTGGAGTGTCTTTCAGCTTCCTTGTCTCTCCCTCAATGCTGGCTTTC 244
 QY 285 CTGGTGTGCACTTCCTATTCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
 DB 245 CTGGTGTGCACTTCCTATTCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 304
 QY 345 AGGATCACAAATGAGATCTTTCTGATCGTATGAGAACTCCACCTCCACAGATTTATC 404
 DB 305 AGGATCACAAATGAGATTTTGTGATGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
 QY 405 AGCTGGCGAGCCAGGTGAAGAGGCGCTGAAAGTGTGTACATGAAGTCCCTGTCTTC 464
 DB 365 AGCTGGCGAGCCAGGTGAAGAGGCGCTGAAAGTGTGTACATGAAGTCCCTGTCTTC 424
 QY 465 GGTCTCTACCAAGAAGTTCGCTGTAAGTGTGTGAGGAGGCGAGTGTCAATGCCCTAC 524
 DB 425 GGTCTCTACCAAGAAGTTCGCTGTAAGTGTGTGAGGAGGCGAGTGTCAATGCCCTAC 484
 QY 525 TACTGTGAGATTCAGATCCCGCCACACCTGGCAGAGAGGTTGATCGGCCCATGGT 584
 DB 485 TACTGTGAGATTCAGATCCCGCCACACCTGGTGGAGGAGCGCGCTCATGGCC 544
 QY 585 GTGAGAGAGTTGTAAATTTGGCCACCCCGAGCACTGGGCACTGAAATCTTCTGTGTAA 644
 DB 545 GAGGAGCGCGTAGTCAATGCTGCCCGCGCGCGCTCCCTGAAGTCTCTTGTGTCAAC 604
 QY 645 TCTGTGTGGCTTCCTCCCATTTGACCCCAAGATGCTGCGAGAGACTCAGAGCAACAGCTGC 704
 DB 605 TCAGTGTGGTTCCTCCCGAGCTCCCAAGAGTCAAGAGAGAGAGAGAGAGAGAGAGTGC 664
 QY 705 AGTTTGTGCTGATGCGCCATGCTGTCAGAGTGAACACGCTTCACTACCCCTGGCTTCGCC 764
 DB 665 AGCTTGTGCTGATGCGCCCGCGGTGTGGAGTGTGAGTGTGCTTCAACACCGCTTCCT 724
 QY 765 AACAGTCCCTACCGCGCATGCGCGCTGCGAGTGGTCTCGGGGGGAGCGCCACTCT 824
 DB 725 GACAGCCCTTACCGCGCTCATGCGCGCTGCGAGTGGGCGCTCGCGGGGAGCGCCACTCA 784
 QY 825 GTGCTGAGCTTCACTTCGGAAGCTTTGATGTGCTGCTGCTGATGAGCATGGCAGTGCAC 884
 DB 785 GTGCTGAGCTTCACTTCGGAAGCTTTGATGTGCTGCTGATGAGCATGGCAGTGCAC 844
 QY 885 CTGCTACCGTGTATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 944
 DB 845 CTGCTACCGTGTATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 904
 QY 945 GGCACCTTCTACCTCTCAAACTGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
 DB 905 GGCACCTTCTACCTCTCAAACTGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
 QY 1005 ACGTGTATTAACCAATATCTGACCGGCGCATCTCTGGCTTTGAGGCGCATCTTCTTCAG 1064
 DB 965 ACCTGTATTAACCAATATCTGACCGGCGCATCTCTGGCTTTGAGGCGCATCTTCTTCAG 1024
 QY 1065 CCCAGATGAGCAGTGTGGCGGCTTTTGTAGTGAACACCCCAAGGAGCATTTAGCAGCCCC 1124
 DB 1025 CCTAGGATGAGCAGTGTGGAGGCGGCTTACGTAAAGCCCAAGGAGCATTTCAACAGCCCC 1084
 QY 1125 TACTATCCAGGCGCACTCCCGCCCAACATCACTGCAATGGAATATCAAGGTGCCCAAC 1184

1085	TACTACCCAGGCCACTACCCACCACAAATTGACTGCACATGGAACTATTGAGGTGCCCAAC	1141
Db		
1185	AACCGGAACGTGAAGGTGCGCTTCAAAACTCTTCTATCTGTGTGGACCCAAAGCTACCAAGTG	1244
QY		
1145	AACCAGCATGTGAAGGTGAGCTTCAAAATCTTCTACTCTGTGGAGCCCGGGTGCCTGGC	1204
Db		
1245	GGCTCTTCGACCAAGGACTATGTGGAGATCAACCGGGGAGAAGTACTCGGTGAGAGGTCC	1304
QY		
1205	GGCACTCTGCCCCAAAGACTAGTGGAGATCAATGGGAGAGAAATACTCGGAGAGAGGTCC	1264
Db		
1305	CAGTTTCTGTGTAGCAGCAACAGCAGCAAGATTAAGTCCACATTCCTCCATTTCTGATCACTCG	1364
QY		
1265	CAGTTCTGTGTACCCAGCAACAGCAACAGATCAAGTTCCGTTCCACTCAGATCAAGTCC	1324
Db		
1365	TACACGGACACCGGGTTCCTAGCTGAGTACTCTCTCTACGACTCCAAAGCACCCGTGCCCA	1424
QY		
1325	TACACGACACCGCTTCTTAGCTGATACCTCTCTCTACGACTCCAGTGAACCATGCCCG	1384
Db		
1425	GGGATGTTCAATGTGCAAGACTGAGACGGTGCATTCGAAAGGAAATGCGCTCGGACGGCTGG	1484
QY		
1385	GGGCAGTTCACTGTGCCACCGGGCGGTGTATCCGGAAGAGCTGGCTGTGATGGCTGG	1444
Db		
1485	GCAGACTGCCCGGATTAAGTGAATGAGCGTTACTGCCGATGCATGCCAATGCCACCCACAGTTC	1544
QY		
1445	GCCGACTGCACGACACACAGGATGACTCAACTGCAGTTGCCAGCGCCGACCAAGTTC	1504
Db		
1545	ACGTGCNAAAACCAAGTTCTCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT	1604
QY		
1505	ACGTGCAAGAAACAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT	1564
Db		
1605	GGGACGGAAGTGAACGAGGAGGGTGAGCTGTCTGTCTGGAGTTCAGTGTTCGAAT	1664
QY		
1565	GGAGACAACAGSCACGAGCAGGGGTGCAGTTGTCCGGCCAGACCTTCAAGTGTTCGAAT	1624
Db		
1665	GGGAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCT	1724
QY		
1625	GGGAGTGTCTCTCGAAAGCCAGTAGTGCATATGGGAAGGACGACTGTGGGACGGGTCC	1684
Db		
1725	GACGAGGCTTCATGTGACAGCGTGAATGTCTGTCTTTGCACCAATATACCTACCCGCTGC	1784
QY		
1685	GACGAGGCTCTCTGCCCAAGGTGAACGTCTGTCATTTGTACAAAACACACTTACCGCTGC	1744
Db		
1785	CAAAATGGCTCTGTCTGAGCAAGGGCAACCTGTAGTGTGATGGGAAGACGACTGTAGC	1844
QY		
1745	CTCAATGGGCTCTGTGTTGAGCAAGGGCAACCCCTGAGTGTGACGGAAGGAGGACTGTAGC	1804
Db		
1845	GATGGCTCCGATCAGAAAAACTCTGACTGTGGCTCGGATCCCTTTTACAAAACAGGCTCCG	1904
QY		
1805	GACGGCTCAGTAGAAGAGACTCCGACTGTGGGCTCGGGTCAATTCACGAGACAGGCTCGT	1864
Db		
1905	GTGGTGTGTGGCACGAATGCGACGAGGGCGAGTGGCCCTTGGCAGGTGAGCCTCCACGGC	1964
QY		
1865	GTTGTTGGGGCACGGATGCGGATGAGGGCAGTGGCCCTGGCAGGTAAAGCTTGCATGCT	1924
Db		
1965	CTGGGCCAGGGCCACTGTGTGGGGCTCGCTCATCTCTCTGTGATGGGTGTGCTGTGCA	2024
QY		
1925	CTGGGCCAGGGCCACATCTGCGGTGTTCCTCATCTCTCCAACTGGGTGTCTCTGGC	1984
Db		
2025	GCTCATTTGCTTTTCAGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGC	2084
QY		
1985	GCACATGCTACATCGATGACAGGATTCAGGTACTCAGACCCCAACGAGTGGACGGCC	2044
Db		
2085	TTCTGGGCTGTGTCGACAGACAGACGCGAGTGCCTCTGGGTGACAGGATGGAAGCTC	2144
QY		
2045	TTCTGGGCTTGACAGACAGACGCGAGCGCCCTTGGGTGACAGAGCGCAGGCTC	2104
Db		
2145	AAACGATATCATCAACCAACCTTCTCTCAATGATTTTCACTTTCGACTATGACATCGCCTTG	2204
QY		
2105	AACGCGATCATCTCCCAACCCCTTCTTCAATGACTTTCACCTTTCGACTATGACATCGCGCTG	2164
Db		
2205	CTGGAGCTGAGAGTGGGTGGAGTACAGCACGCTGTGGGCCCACTGTGCTGCTGAT	2264
QY		
2165	CTGGAGCTGGAGAAAACGGCAGAGTACAGTCAATGGTGGGCCCACTGTGCTGCGGAC	2224
Db		

Qy	2265	GTCTACCAATGCTTTCCCTGCTGTCACAAAGGCCATCTGGGTCTCAGCGCTGGGGGCACACAAAA	232
Db	2225	GCCTCCCAATGCTTTCCCTGCTGCGGCAAGGCCATCTGGGTACAGGGCTGGGACACACCCAG	2284
Qy	2325	GAGGGAGGTACCGGAGCGCTGATCTTCGACGAAGGTTGAGATCCGTGTCTATCAACCCAGACC	2384
Db	2285	TATGAGAGCACTGGCGCGCTGATCTCTGCAAAAGGGTGAGATCCGCGTCATCAACCCAGACC	2344
Qy	2385	ACCTGTGAGGACCTCATCGCGCAGCAGATCACCCACGAATGATGTGTGGGTTCCTTC	2444
Db	2345	ACCTGCGAGAACCTCTCGCGCAGCAGATCACGCGCGCATGATGTGCGTGGGCTTCCTC	2404
Qy	2445	AGTGGGGGTGTGACTCTCTGCGAGGTGACTCTGTGTGGGCCCTTGTCAAGCGCGGAGAAA	2504
Db	2405	AGCGGCGGGTGGACTCTCTGCCAGGGTGATTCGCGGGGACCCCTGTCCAGCGTGGAGGCG	2464
Qy	2505	GATGGCGGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC	2564
Db	2465	GATGGCGGATCTTCAGGCGGTTGTGGTAGCTGGGGAGACGGCTGCGCTCAGAGGAAC	2524
Qy	2565	AAGCCAGGCGTGTACAAAGGTCCTCTGTAGTTCGGGACTTGGATCAAGAGCACACTGGG	2624
Db	2525	AAGCCAGGCGTGTACAAAGGTCCTCTCTGTTTCGGGACTGGATCAAGAGAACACTGGG	2584
Qy	2625	GTATACAGCATGGACAGACAGACGCCGACCAAAACACCCACAGGGATGCCGACATGCACA	2684
Db	2585	GTATAGGGGCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCATGTGTCACC	2641
Qy	2685	CCTGGATACAGAGAGGAACACTGACGACATTTATGCTGTGGCTCTCCCGCCCAACACA	2744
Db	2642	CCAGTGTGCAAG-CCTGCGAGCTGGAGACTGACCGCTGACTGCACACAGGGGCC-CCAGA	2699
Qy	2745	ACCAGACTGTGAATGTGCATCTCTAGGACTCAGAT	2780
Db	2700	ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT	2735

RESULT 12

ABZ58501

ID

ABZ58501 standard; cDNA; 3147 BP.

XX

ABZ58501;

XX

13-MAY-2003

(first entry)

DE

Transmembrane serine protease 1 (MTSP1) protein domain cDNA.

XX

Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;

KW

cytostatic; dermatological; cardiant; vulnerary; ophthalmological;

KW

gene therapy; gene; ss.

XX

Homo sapiens.

OS

XX

Key

Location/Qualifiers

XX

1865..2590

FT

/*tag= a

FT

/product= "MTSP1 protease domain"

XX

WO2003004681-A2.

PN

XX

16-JAN-2003.

XX

03-JUL-2002; 2002WO-US21208.

PF

XX

03-JUL-2001; 2001US-302939P.

PR

XX

(CORV-) CORVAS INT INC.

PA

XX

Madison EL, Ong EO;

PI

XX

WPI; 2003-239207/23.

DR

P-PSDB; ABP72377.

DR

XX New type-II membrane-type serine protease 20 polypeptides, useful for
PT preparing a medicament for diagnosing, treating or preventing cancer,
PT dermatological disorders, aberrant wound repairs or crest syndromes -
XX Disclosure; Page 200-202; 216pp; English.

XX The present sequence is that of cDNA encoding the protease domain
CC of human type II transmembrane serine protease 1 (MTSP1). The
CC invention relates to novel human type II transmembrane serine protein
CC 20 (MTSP20) (see ABP72374) and nucleic acids encoding it (see
CC AB258499). Also claimed are methods of inhibiting tumour initiation,
CC growth or progression by inhibiting MTSP20 activity, and of treating
CC or preventing a disease or disorder associated with undesired and/or
CC uncontrolled angiogenesis or neovascularisation, especially
CC undesired angiogenesis associated with solid neoplasms, vascular
CC malformations and cardiovascular disorders, chronic inflammatory
CC diseases, aberrant wound repairs, circulatory disorders, crest
CC syndromes, dermatological disorders and ocular disorders using an
XX inhibitor of MTSP20.

XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGACCGCCAAACCATGGTAGCAATCGGGCCGCAAGCGCGAGGGGGCTCTCAG 104
DB 5 GAGCGGCTCGGGGTACCATCGGGAGCGATCGGGCCCGCAGGGGGGCGCGAG 64

QY 105 GACTTCGGCGGGGACTCAAGTACACTCCCGGTAGAGACATGAATGGCTTTAGAGAG 164
DB 65 GACTTCGGCGGGGACTCAAGTACACTCCCGGTAGAGACATGAATGGCTTTAGAGAG 124

QY 165 GGTGTGGAGTTCTCTGCTCGTCAAAATGCCAAGAGTGGAGAGCGAGGCCCGCAGGCG 224
DB 125 GGCCTGGAGTTCTCTGCTCGTCAAAATGCCAAGAGTGGAGAGCGAGGCCCGCAGGCG 184

QY 225 TGGGTGGTGGTGGGAGTCTGTTCAGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284
DB 185 TGGGTGGTGGTGGGAGTCTGTTCAGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244

QY 285 CTGCTGGGACTTCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
DB 245 CTGCTGGGACTTCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 304

QY 345 AGGATCAAAATGAGATCTTTCTCGATGCGTATGAGAACTCCACTCCACAGAGTTTATC 404
DB 305 AGGATCAAAATGAGATTTTGTGGATGCGCTGAAGCTGTGTACAGAGTCCCACTCC 364

QY 405 AGCTGGCCAGCGAGGTGAAGGAGCGCTGAAGCTGTGTACATGAAGTCCCTGTCTCTG 464
DB 365 AGCTGGCCAGCGAGGTGAAGGAGCGCTGAAGCTGTGTACAGAGTCCCACTCC 424

QY 465 GGTCTCTACACAGAGTGGCTGTAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 524
DB 425 GGTCTCTACACAGAGTGGCTGTAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 484

QY 525 TACTGTGTACAGTTTACATCTCCCGACACTGCGCAGAGAGTTGATCGGCCATGGCT 584
DB 485 TACTGTGTACAGTTTACATCTCCCGACACTGCGCAGAGAGTTGATCGGCCATGGCT 544

QY 585 GTGAGCGAGTTGTAAATTGCCACCCGAGCAGCGGCACTGAAATCTTCTGTCTTAAACA 644
DB 545 GAGGAGCGCGTAGTGTATGCTGCCCGCGGGCGCTCCCTGTAAGTCTTTGTGGTCAAC 604

QY 645 TCTGTGTGGCTTCCCGATGAGTCCCGAGAAATGTGTGAGAGAGTCAAGACACAGCTGC 704
DB 605 TCAGTGTGGCTTCTCCCGAGCGAGTCTCAAAACAGTACAGAGGACCCAGGACACAGCTGC 664

QY 705 AGTTTTCCTGTCATGCCCATGGTGTGACAGTACACGCTTCACTACCTCGCTCTCC 764

DB 665 AGCTTTGGCTGCACGCGCCGCGGTGTGAGAGTGTGCGCTTCAACACGCGCCGCTTCCCT 724

QY 765 AACAGTCCCTACCGCGGCATGCGCGTGCAGTGGTCTTGGGGGGGACGCCGACTCT 824

DB 725 GACAGCCCTTACCGCGGTCTCATGCGCGTGCAGTGGGCGCTTGGGGGGGACGCCGACTCA 784

QY 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTGCTCTGTGATGAGCATGGCAGTGAC 884

DB 785 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTGCTCTGTGATGAGCATGGCAGTGAC 844

QY 885 CTGCTGACCGTGTATGATGAGCCTGAGCCCAATGGAACCCCACTGTGTGTGGGTGTGT 944

DB 845 CTGCTGACCGTGTATGATGAGCCTGAGCCCAATGGAACCCCACTGTGTGTGGGTGTGT 904

QY 945 GGCACCTTCTCACCTCTCAACCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004

DB 905 GGCACCTTCTCACCTCTCAACCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964

QY 1005 ACGTGTATAACCAATACTGACCGGCGCATCTCTGCTTTGAGGCGCATCTTCTTCCAGCTG 1064

DB 965 ACGTGTATAACCAATACTGACCGGCGCATCTCTGCTTTGAGGCGCATCTTCTTCCAGCTG 1024

QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCCAAGGAGACATTTAGCAGCCCC 1124

DB 1025 CCTAGGATGAGCAGCTGTGGAGGCGCTTACGTAAAGCCCAAGGCGCATTTCAACAGCCCC 1084

QY 1125 TACTATCCAGGCACTACCGCCCAACATCACTGACATGCACTGGATATCAAGGTGCGCAAC 1184

DB 1085 TACTATCCAGGCACTACCGCCCAACATTTGACTGACATGGAACATTTAGGTGCGCAAC 1144

QY 1185 AACCGAAGCTGAAGTGGCTTCAAACTCTTCTTATCTGTGTGACCCCAAGTCCAGTG 1244

DB 1145 AACCGAAGCTGAAGTGGCTTCAAACTCTTCTTATCTGTGTGACCCCAAGTCCAGTG 1204

QY 1245 GGTCTCTGACCAAGGACTATGTGGAGTCAACCGGGAGAGTACTGCGGTGAGAGTTC 1304

DB 1205 GGCACCTGCCCCAAGGACTACGTGGAGTCAATGGGGAGAAATACTCGGAGAGAGTTC 1264

QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATACAGTCCACTTCCATTTCTGTACTCTG 1364

DB 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATACAGTCCACTTCCACTCAGATCAGTTC 1324

QY 1365 TACACGACACCGGGTTCTTAGTGTAGTACTCTCTTACACTCCAAAGACCCGCTGCCCA 1424

DB 1325 TACACGACACCGGGTTCTTAGTGTAGTACTCTCTTACACTCCAGTCCAGTACCCG 1384

QY 1425 GGGATGTGTATGTGCAAGACTGACCGTGTGATCGAAAGAACTGCGCTGCGAGCGGTGG 1484

DB 1385 GGGCAGTTTCACTGCGCGCACCGGGCGGTGTATCGGAAAGAGTGTGCTGTGTGGTGG 1444

QY 1485 GCAGACTGCGCGGATTTATGTATGATGAGCGTTTACTGCGGATGCAATGCCACCCAGCTTC 1544

DB 1445 GCGCAGTGTACCGACACAGCGATGAGTCACTGTGAGTTGCGACGCGCGCCACAGTTC 1504

QY 1545 ACGTCAAAACACAGTTTGTCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604

DB 1505 ACGTCAAAACACAGTTTGTCAAGCCCTCTTCTGGGTCTGTGACAGTGTGAAGACTGC 1564

QY 1605 GGGGACGAGTGTACAGGAGGGCTGCGAGTGTCTCTGCTGGGAGTTTCAAGTGTTCAT 1664

DB 1565 GGGGACGAGTGTACAGGAGGGGTGCGAGTGTCTGCGCGCCAGACCTTCAAGTGTTCAT 1624

QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAGAGTGTAAATGGGAAGGACAACTGTGGAGATGGTCT 1724

DB 1625 GGGAGTGTCTCTGAAAGCCAGAGTGCATGGAAGAGACGACTGTGGGGAGCGGTTC 1684

QY 1725 GACGAGCTTATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGCTGC 1784

DB 1685 GACGAGCTTCTCTGCCCCAAGTGAAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1744

QY 1785 CAAATGTGCTGTGTGAGCAAGGCGCAACCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGT 1844

DB 1745 CTAATGTGCTGTGTGAGCAAGGCGCAACCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGT 1804

```
QY 1845 GATGCTCCGATGAGAAACACTGTGACTGTGGGCTCGGATCCTTTACCAACAGGCTCGC 1904
D 1805 GACGCTCAGATGAGAACTGCGACTGTGGCTGGGTCTATTACAGAGACAGGCTCGT 1864
QY 1905 GTGGTGGTGGCAGATCGGAGCGAGGCGAGTGGCCCTGGCAGGTGAGCTCCAGGCC 1964
D 1865 GTTGTGGGGGCAAGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCAGGT 1924
QY 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCCTCGCTCATCTCTCTGACTGCTGGTCTCTGCA 2024
D 1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGCC 1984
QY 2025 GCTCATTTGCTTTAGGATGACAAAATTTCAAGTACTCAGACTACAGATGTGGACGGCC 2084
D 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAAGCTGAGCGGCC 2044
QY 2085 TTCTGGGTCTGTGGACACAGAGCAAGCGAGTGCCTCTGCTGCTGCTGGTGGTGGAGCTGAAGCTC 2144
D 2045 TTCTGGGTCTGGACACAGAGCAAGCGAGTGCCTCTGCTGCTGCTGGTGGTGGAGCTGAAGCTC 2104
QY 2145 AAACGTATCATACCCACCCCTTCTTCAATGATTTCACTTCGACTATGATCATCGCCTTG 2204
D 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTCGACTATGATCATCGCCTTG 2164
QY 2205 CTGGAGCTGGAGAGTGGTGGAGTACAGCACCGTGTGGCCGCCATCTGCTGGCTGTGAT 2264
D 2165 CTGGAGCTGGAGAGTGGTGGAGTACAGCACCGTGTGGCCGCCATCTGCTGGCTGTGAT 2224
QY 2265 GCTACCCATGCTTCTCCCTGTGGCAAGCGCATCTGGGTGACAGCTGGGGGACACAAAA 2324
D 2225 GCTTCCCATGCTTCTCCCTGTGGCAAGCGCATCTGGGTGACAGCTGGGGGACACCCAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTCAGAGGCTGAGATCCGTGTCATCAACAGAC 2384
D 2285 TATGAGGCACTGGCGGCTGATCTCTCAAAGGCTGAGATCCGTGTCATCAACAGAC 2344
QY 2385 ACCTGTGAGACCTCATCGCGAGCAGATCAACCCAGAGATGATGTGTGGTTCCTC 2444
D 2345 ACCTGTGAGACCTCATCGCGAGCAGATCAACCCAGAGATGATGTGTGGTTCCTC 2404
QY 2445 AGTGGGGGTGTGACTCTCTCCAGGGTACTCTGTGGGCCCTTGTCAAGCGGAGAAA 2504
D 2405 AGCGGGCGGTGACTCTCTCCAGGGTACTCTGTGGGGGACCCCTGTCCAGCGTGGAGG 2464
QY 2505 GATGGCGGAATGTTCCAGGCTGTGTGGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGAAC 2564
D 2465 GATGGCGGAATGTTCCAGGCGGTGTGTGGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGAAC 2524
QY 2565 AAGCCAGGCGGTGACACAGGCTCTCTGTAGTTGCGGACTGGATCAAGAGCACTGGG 2624
D 2525 AAGCCAGGCGGTGACACAGGCTCTCTGTAGTTGCGGACTGGATCAAGAGCACTGGG 2584
QY 2625 GTATAGCAGATGACAGACAGCGGACCAACACACAGGATGCGCGACATGCACA 2684
D 2585 GTATAGGCGCGG---GCCACCAATGTGTACCTGGGGGACCACTGCTCAC 2641
QY 2685 CTTGGATACAGGAGAGAACATGTGACAGATTTATGTGTGGCTCCCCCCCCCAACA 2744
D 2642 CCAGTGTGACG---CTTGGAGCTGGAGACTGGACCGCTGACTGCACCGAGGCC---CCAGA 2699
QY 2745 ACCGAGCTGTGAATGATCTCTTTAGGACTCAGAGT 2780
D 2700 ACATACATGTGACTCATCTCCAGGCTCCAAAT 2735
```

RESULT 13

ID ABZ22450

XX ABZ22450 standard; cDNA; 3147 BP.

AC ABZ22450;

XX

PT 24-MAR-2003 (first entry)

XX DB Human membrane-type serine protease MTSPl encoding cDNA SEQ ID NO:1.
XX KW Human; membrane-type serine protease; enzyme; MTSPl0; cytosolic;
KW type-II membrane-type serine protease; neoplastic disease; tumour;
KW MTSPl; matriptase; gene; ss.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX CDS 23..2590
XX FT /*tag= a
XX FT /product= "membrane-type serine protease MTSPl"
XX FT /note= "also known as matriptase"
XX PN W0200292841-A2.
XX PD 21-NOV-2002.
XX PF 14-MAY-2002; 2002WO-US15332.
XX PR 14-MAY-2001; 2001US-291001P.
XX PA (CORV-) CORVAS INT INC.
XX PI Madison EL, Yeh J;
XX DR WPI; 2003-129309/12.
XX DR P-PSDB; ABP56619.

New polypeptides comprising the protease domain of a type-II membrane-type serine protease (MTSPl0), or its mutants, useful for diagnosing neoplasms or malignancies, or for screening for MTSPl0 inhibitors for treating such diseases

Example 1; Page 177-181; 198pp; English.

The present invention describes a polypeptide comprising a purified single or two chain polypeptide, which comprises the protease domain of a type-II membrane-type serine protease (MTSPl0) or its catalytically active portion, or a mutant of it, where up to 50 % of the amino acids are replaced with another amino acid, and the resulting polypeptide is a single chain or two chain polypeptide that has a catalytic activity of at least 1-10 % of the unmutated polypeptide. MTSPl0 has cytostatic activity. The polypeptide containing the protease domain of the MTSPl0 is useful for detecting a neoplastic disease, and for diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, or monitoring tumour (e.g. breast, cervix, prostate, lung, ovary or colon tumour) progression and/or therapeutic effectiveness. An inhibitor of the polypeptide containing the protease domain of MTSPl0 is useful for treating or preventing neoplastic disease in a mammal. An inhibitor of the activation cleavage of the zymogen form of the MTSPl0 polypeptide is useful for inhibiting tumour initiation, growth or progression, or treating (pre-)malignant conditions of the e.g. breast, cervix, prostate, lung, ovary or colon. The present sequence encodes human MTSPl (also known as matriptase), which is used in an example from the present invention.

Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGAGCCGCCAAACCATGGTAGCATCGGGCCCGCAGGCGGGGGCTCTAG 104
D 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAGGCGGGGGCGCGAG 64
QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAACATCAATGGCTTTTCAGGAG 164
D 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAACATCAATGGCTTTTCAGGAG 124
QY 165 GGTGTGGAGTTCTCTCGCTGGCAACAATGCCAAGAAAGTGGAGAGCGAGGCCCGCGGC 224

QY 2385 ACCTGTGAGGACCTCATCGCCGAGCAGATCACCCACGAATGATGTGTGGTTTCTC 2444
 DB |||||
 QY 2345 ACCTGCGAAGACCTCTCTGCGCAGCAGATACGCGCGCATGATGTGTGGTTTCTC 2404
 DB |||||
 QY 2445 AGTGGGGTGTGACTCTCTGCGCAGGTCATCTGTGTGGCCCTTGTCAAGCGCGAGAAA 2504
 DB |||||
 QY 2405 AGCGCGCGGTGGACTCTCTGCGCAGGTCATCTGTGTGGCCCTTGTCAAGCGCGAGAAA 2464
 DB |||||
 QY 2505 GATGGGCAATGTTTCCAGCGTGTGTGTGAGCTGGGTGAAGCTGGCTCAGAGAAC 2564
 DB |||||
 QY 2465 GATGGCGGATCTTCCAGCGGTGTGTGTGAGCTGGGGAGACGGTGGCTCAGAGAAC 2524
 DB |||||
 QY 2565 AAGCAGCGGTGTACAAAGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACTGGG 2624
 DB |||||
 QY 2525 AAGCAGCGGTGTACAAAGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACTGGG 2584
 DB |||||
 QY 2625 GTATAGCAGCATGCACAGACCGCACCAACACACACAGGATGCCGACATGCACA 2684
 DB |||||
 QY 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641
 DB |||||
 QY 2685 CCTGGATACAGGAGGAGAACACTGACGACATTTATGTGTGGCTTCCCGCCCAACACA 2744
 DB |||||
 QY 2642 CCAGTGTGACG-CCTGCAGCTGGAGACTGGACGCTGACTGCACGAGCGCC-CCAGA 2699
 DB |||||
 QY 2745 ACCCAGACTGTGAATCTGATCCTTTAGGACTCAGAGT 2780
 DB |||||
 QY 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735
 DB |||||

RESULT 14

ABZ22451
 ID ABZ22451 standard; cDNA; 3147 BP.

XX AC ABZ22451;

XX DT 24-MAR-2003 (first entry)

XX DE Human MTSP1 protease domain encoding cDNA SEQ ID NO:3.

XX KW Human; membrane-type serine protease; enzyme; MTSP10; cytostatic;

XX KW type-II membrane-type serine protease; neoplastic disease; tumour;

XX KW MTSP1; matriptase; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 1865..2590

XX FT /*tag= a

XX FT /partial

XX FT /product= "MTSP1 protease domain"

XX FT /note= "no start codon given"

XX XX WO200292841-A2.

XX XX 21-NOV-2002.

XX XX 14-MAY-2002; 2002WO-US15332.

XX XX 14-MAY-2001; 2001US-291001P.

XX XX (CORV-) CORVAS INT INC.

XX PA Madison EL, Yeh J;

XX PI WPI; 2003-129309/12.

XX DR P-PSDB; ABP56620.

XX XX New polypeptides comprising the protease domain of a type-II

XX PT membrane-type serine protease (MTSP10), or its mutants, useful for

XX PT diagnosing neoplasms or malignancies, or for screening for MTSP10

XX PT inhibitors for treating such diseases

XX XX

Example 1; Page 183-185; 198pp; English.

The present invention describes a polypeptide comprising a purified single or two chain polypeptide, which comprises the protease domain of a type-II membrane-type serine protease (MTSP10) or its catalytically active portion, or a mutin of it, where up to 50 % of the amino acids are replaced with another amino acid, and the resulting polypeptide is a single chain or two chain polypeptide that has a catalytic activity of at least 1-10 % of the unmutated polypeptide. MTSP10 has cytosolic activity. The polypeptide containing the protease domain of the MTSP10 is useful for detecting a neoplastic disease, and for diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, or monitoring tumour (e.g. breast, cervix, prostate, lung, ovary or colon tumour) progression and/or therapeutic effectiveness. An inhibitor of the polypeptide containing the protease domain of MTSP10 is useful for treating or preventing neoplastic disease in a mammal. An inhibitor of the activation cleavage of the zymogen form of the MTSP10 polypeptide is useful for inhibiting tumour initiation, growth or progression, or treating (pre-)malignant conditions of the e.g. breast, cervix, prostate, lung, ovary or colon. The present sequence encodes the protease domain of human MTSP1 (also known as matriptase), which is used in an example from the present invention.

Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;

Best Local Similarity 81.2%; Pred. No. 0;

Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGGCGCGCAAGCGCGGGGCTCTCAG 104
 DB |||||
 QY 5 GAGCGGCTCTCGGGTACCATGGGGAGCATCGGCGCGCAAGCGGGGCGGCGGCGGAG 64
 DB |||||
 QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGCTTTGAGGAG 164
 DB |||||
 QY 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGCTTTGAGGAA 124
 DB |||||
 QY 165 GGTGTGGAGTCTCTGCTCGGAACTGCAAGAAAGTGAGAGAGCGGCCCGCGCC 224
 DB |||||
 QY 125 GGGTGGAGTCTCTGCGAGTCAACACAGTCAAGAGGTGAAAGCATGGCGCGGCGC 184
 DB |||||
 QY 225 TGGGTGGTGTGGTGGCAGTGTGTTGAGTCTCTCTTGTCTCTCCCTCATGGTGGCTTG 284
 DB |||||
 QY 185 TGGGTGGTGTGGCAGCGCTGTGATCGGCTCTCTTGTCTCTGTGGGATCGGCTTC 244
 DB |||||
 QY 285 CTGGTGTGGCATTCCATTATCGGAATGCGGGTTCAAAAAGTCTTCAATGGCCCATCTG 344
 DB |||||
 QY 245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTTCTTCAATGGCTACATG 304
 DB |||||
 QY 345 AGGATCACAATGAGATCTTCTTGGATGCTATGAGAACTCCACCTCCACAGAGTTTATC 404
 DB |||||
 QY 305 AGGATCACAATGAGAAATTTTGTGGATGCTACGAGAACTCCAACTCCACTGAGTTTGA 364
 DB |||||
 QY 405 AGCTGGCCAGCCAGGTGAAGAGCGCTGAGTGTCTGTACAATGAAGTCCCTGTCCTG 464
 DB |||||
 QY 365 AGCTGGCCAGCAAGGTGAAGAGCGCTGAGTGTCTGTACAAGTCTCCATTCCTG 424
 DB |||||
 QY 465 GGTCCCTACCAAGAAGTCTGTAAGTGTCTGAGTGTGAGGCGAGTGTCTATCGCTTAC 524
 DB |||||
 QY 425 GGGCCCTACCAAGAGTCTGAGTGTGAGGCTTTCAGGCGGAGCGTCTATCGGCTTAC 484
 DB |||||
 QY 525 TACTGTGAGAGTTTCTGAGTCTCCCGCCACCTGGGAGAGAGGTTGATCGCGCATGGCT 584
 DB |||||
 QY 485 TACTGTGAGTTTCTGAGTCTCCCGCAGCACCCTGTTGGAGAGCGCGCTCATGGCC 544
 DB |||||
 QY 585 GTGGAGCGAGTTGTAACTATGTCACCCGAGCACCGGCGACTGAAATCTCTTCTGTAAACA 644
 DB |||||
 QY 545 GAGGAGCGGTAGTCTATGTCGCCCGCGCGCGCTCCCTGAAGTCTCTTGTGGTCCACC 604
 DB |||||
 QY 645 TCTGTGGTGGCTTCCCAATGACCCCGAGAGTGTGCGAGAGCACTTCAGGACAAACAGCTGC 704
 DB |||||
 QY 605 TCAGTGGTGGCTTCTCCCGCAGGACTCCAAAACAGTACAGAGAGACCCAGACAAACAGCTGC 664
 DB |||||

QY 705 AGTTTTCCTGATGCCATGCTGAGAGTGAACAGCTTCACTACCCCTGGCTTCCC 764
Db 665 AGCTTTGGCTTGACAGCCCGCGGTGTGAGCTGATGCGCTTACACAGCCCGCGCTTCCCT 724
QY 765 AACAGTCCCTACCCCGCGCATCCCGCTGCCAGTGGTCTCGCGGGGAGCCGACCTCT 824
Db 725 GACAGCCCTACCCCGCTCATGCCCGCTGCCAGTGGTCTCGCGGGGAGCCGACCTCA 784
QY 825 GTGCTGAGCTCACCTTCGGAAGCTTTGATGCTGCTTGTGATGAGCATGGCAGTGC 884
Db 785 GTGCTGAGCTCACCTTCGCGAGCTTTGACCTTGGCTCTGCGAGAGCGCGGAGCGAC 844
QY 885 CTGCTCACCGTGTATGATGCTGAGCCCTGAGAACCCGCTGCTGCTGCTGCTGCTGCT 944
Db 845 CTGCTGAGCGGTGTAACAACCCCTGAGCCCTGAGGAGCCCGCTGCTGCTGCTGCTGCT 904
QY 945 GGCACCTTCTCACCCCTCTCAACACCTGACTTTCCTCTCTCCAGAACGCTTCTCTCTG 1004
Db 905 GGCACCTTACCTCTCTCAACCTGACTTTCCTCTCTCCAGAACGCTTCTCTCTCACT 964
QY 1005 AGCTGTATTAACAATGACCGCGGCAATCCTGCTTTGAGGCGACCTTCTTCCAGCTG 1064
Db 965 ACATGTATAACCAACATGAGCGCGGCAATCCTGCTTTGAGGCGACCTTCTTCCAGCTG 1024
QY 1065 CCCAGATGAGCAGCTGCGCGCTTTTGTAGTGACCCCAAGGAGATTTAGAGCGCCC 1124
Db 1025 CTTAGATGAGCAGCTGCGCGCTTTTGTAGTGACCCCAAGGAGATTTAGAGCGCCC 1084
QY 1125 TACTATCCAGGCGCACTACCCCGCCAACTGACATGCAATGCAATATCAAGGTGCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCCCGCCAACTGACATGCAATGCAATATCAAGGTGCCAAC 1144
QY 1185 AACCGAAGCTGAGGTGCGCTTCAAACTTCTTATCTGTTGAGCCCAAGTCCAGTGT 1244
Db 1145 AACCGAAGCTGAGGTGCGCTTCAAACTTCTTATCTGTTGAGCCCAAGTCCAGTGT 1204
QY 1245 GGCCTCTGACCAAGGACTATGAGTATCAACGGGAGAGTACTGCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTATGAGTATCAACGGGAGAGTACTGCGGTGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATCACTCG 1364
Db 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATCACTCG 1324
QY 1365 TACAGCAGACCGGGTCTTACGTGAGTACTCTCTACGATCAAGTCCAGTCCAGTCC 1424
Db 1325 TACAGCAGACCGGGTCTTACGTGAGTACTCTCTACGATCAAGTCCAGTCCAGTCC 1384
QY 1425 GGGATGTTTCATGTCAAGACTGAGCGGTGATCCGAAAGGAACTGCGCTGCGAGCGTGG 1484
Db 1385 GGGCAGTTACGTCCGCAAGCGGGCGGTGATCCGAAAGGAACTGCGCTGCGAGTGG 1444
QY 1485 GCAGACTGCCCGGATTAAGTATGAGCGTTTACTGCGGATGCAATGCCACCCAGTTC 1544
Db 1445 GCCGACTGCAACCGACCAAGGATGAGTCAACTCAAGTTCAGTTCAGTCCAGTTC 1504
QY 1545 ACGTGCAAAACAGTCTTCAAGCCCTCTTCTGCGGTGCTGAGTCAAGTCAAGTCTGT 1604
Db 1505 ACGTGCAAAACAGTCTTCAAGCCCTCTTCTGCGGTGCTGAGTCAAGTCAAGTCTGT 1564
QY 1605 GGGCAGGAACTGAGGAGGCGGTGAGTCTGCTGCTGCGGAGTTCAGTGTTCAT 1664
Db 1565 GGGCAGGAACTGAGGAGGCGGTGAGTCTGCTGCTGCGGAGTTCAGTGTTCAT 1624
QY 1665 GGGAAAGTGTCTTCTTCCAGCAGCAAGTGTAAATGGGAGGAGCACTGTGGAGTGGTCT 1724
Db 1625 GGGAAAGTGTCTTCTTCCAGCAGCAAGTGTAAATGGGAGGAGCACTGTGGAGTGGTCT 1684
QY 1725 GACGAGCTTCACTGACGCGTGAATGCTGCTTCTTGTGACCAAAATATACCTACCGTGC 1784
Db 1685 GACGAGCTTCACTGACGCGTGAATGCTGCTTCTTGTGACCAAAATATACCTACCGTGC 1744
QY 1785 CAAAATGGCTCTCTCTGAGCAAGGGCAACCTTGAAGTGTGATGGAGAGCGGACTGTAGC 1844

RESULT 15
AAD47180
ID AAD47180 standard; DNA; 3147 BP.
XX
AC AAD47180;

Db 1745 CTCATAGGCTCTGCTTGGCAAGGGCAACCTGAGTGTGACGGGAGGAGACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACCTGTGACTGTGGGTGCGATCTCTTTACAAACAGAGCTGCG 1904
Db 1805 GACGGCTCAGATGAGAGGACTGTGCGACTGTGGGTGCGGTCAATTCACGAGACAGGCTGCT 1864
QY 1905 GTGGTTGTGTCAGCAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db 1865 GTTGTGGGGGCAACGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCTGCTGCT 1924
QY 1965 CTGGCCAGGGCCACTTGTGTGGGGCCCTGCTCATCTCTCTGACTGTGCTGCTCTGCA 2024
Db 1925 CTGGCCAGGGCCACTTGTGTGGGGCCCTGCTCATCTCTCTGACTGTGCTGCTGCTG 1984
QY 2025 GCTCAATGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGCATGTGAGCGCC 2084
Db 1985 GCACATGCTATCATGATGACAGAGGATTTCAAGTACTCAGACTACAGCATGTGAGCGCC 2044
QY 2085 TTCTGGGCTCTGCTGGACCAAGCGCAGTGCCTCTGGGGTGCAGAGCTGAAGCTC 2144
Db 2045 TTCTGGGCTCTGCTGGACCAAGCGCAGTGCCTCTGGGGTGCAGAGCTGAAGCTC 2104
QY 2145 AAACGTATCATCAACCCCTTCTTCAATGATTTCACTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTCGACTATGACATCGCCTG 2164
QY 2205 CTGAGCTGTGAGAGTCTGCTGAGTACAGCACCTGCTGCGCCCTGCTGCTGCTGCTG 2264
Db 2165 CTGAGCTGTGAGAGTCTGCTGAGTACAGCACCTGCTGCGCCCTGCTGCTGCTGCTG 2224
QY 2265 GCTACCATGCTTCTTCTGCTGCGCAAGGCCATCTGGGTCAAGGCTGGGGGACACAAA 2324
Db 2225 GCTTCCCATGCTTCTTCTGCTGCGCAAGGCCATCTGGGTCAAGGCTGGGGGACACAAA 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTTCCGAGAGGTGAGTCCGTGCTATCAACCCAGACC 2384
Db 2285 TATGAGGCACTTGGCGCGCTGATCTTCCGAGAGGTGAGTCCGTGCTATCAACCCAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGCGAGCAGATCAACCCGAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGGACCTCATGCGCGAGCAGATCAACCCGAGATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGAGCTCTTCCGAGGGTGAATCTGCTGGGGCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGCGCGGTGGACTCTTCCGAGGGTGAATCTGCGGGGACCCCTGTCCAGCTGAGGCG 2464
QY 2505 GATGGGCGAATGTTCCAGGCTGTGTGGTGTGAGCTGGGTGAAGGCTGGCTCAGAGGAC 2564
Db 2465 GATGGGCGAATGTTCCAGGCTGTGTGGTGTGAGCTGGGTGAAGGCTGGCTCAGAGGAC 2524
QY 2565 AAGCCAGGCGGTGTACACAAGGCTCCCTGTAGTTTGGGACTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGGCGGTGTACACAAGGCTCCCTGTAGTTTGGGACTGGATCAAGAGCACTGGG 2584
QY 2625 GTATAGCAGATGGACAGCAGCCGACCAAAACACCCACAGGGATGCCCGCATGACACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCTGTCCACC 2641
QY 2685 CCTGGATACAGGAGAGGAACTGACGACATTTATGCTGTGGGCTTCCCCCCCCCAACACA 2744
Db 2642 CCAGTGTGACG---CCTGAGGCTGGAGCTGGACCGCTGACTGCACCGCGCC---CCAGA 2699
QY 2745 ACCCAGACTGTGAATGCTATCTTAGSACTTCAGAT 2780
Db 2700 ACATCACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

XX DT 24-FEB-2003 (first entry)
XX DE Human membrane-type serine protease, MTSPl DNA.
XX KW Human; serine protease 14; CVSP14; cancer; malignancy; breast; colon;
KW gene therapy; cytostatic; membrane-type serine protease; enzyme; MTSPl;
XX OS gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 23..2590
XX FT /*tag= a
XX FT /product= "Human MTSPl protein"
XX WO200277263-A2.
XX PD 03-OCT-2002.
XX PF 20-MAR-2002; 2002WO-US09039.
XX PR 22-MAR-2001; 2001US-278166P.
XX PA (CORV-) CORVAS INT INC.
XX PI Madison EL, Yeh J;
XX DR WPI; 2003-018939/01.
XX DR P-PSDB; AAE29791.
XX PT New purified CVSP14 polypeptide and encoding nucleic acid molecule,
XX PT useful for diagnosing, preventing and/or treating disorders, such as
XX PT cancers and malignancies of the breast, cervix, prostate, lung, ovary
XX PT or colon -
XX PS Disclosure; Page 167-171; 185pp; English.
XX CC The invention relates to transmembrane serine protease 14 (CVSP14), its
XX CC nucleic acid sequence and the method based on them. The methods and
XX CC compositions of the invention are useful for diagnosing, preventing
XX CC and/or treating conditions associated with the aberrant expression or
XX CC activity of the CVSP14 polypeptide, such as cancers and malignancies of
XX CC the breast, cervix, prostate, lung, ovary or colon. The methods are
XX CC also useful for identifying compounds that will modulate the protease
XX CC activity of CVSP14 polypeptide, and monitoring tumour progression and/or
XX CC therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present
XX CC sequence is human membrane-type serine protease, MTSPl DNA.
XX SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGCCGCAAGCGCGAGGGGCTCTCAG 104
DB 5 GAGCGGCTCGGGTACCATGGGAGGAGATCGGGCCCGCAAGGGCGAGGGGCGCCGAAG 64
QY 105 GACTTCGGCGGGGACTCAAGTACAATCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
DB 65 GACTTCGGCGGGGACTCAAGTACAATCCCGGCAAGAAAGTGAATGCTTTGGAGAA 124
QY 165 GGTGTGGAGTTCCTGCTCGGACAAATGCCAAGAAAGTGGAGAAAGGAGCGCCAGCGC 224
DB 125 GGGTGGAGTTCCTGCGAGTCAACAAAGTGAAGAGTGAAGAGTGGCGCGCGCGC 184
QY 225 TGGTGTGTGTGGGAGTGTCTTCAGTCTCCTTTGCTCTCCCTCATGGCTGGCTTG 284
DB 185 TGGTGTGTGTGGGAGCGGTGCTGATCGGCTCTCTTGTGTGTGGGATCGGCTTC 244
QY 285 CTGGTGTGGCACTTCCATTATCGGAATGTGGGCTTCAAAAAGTCTTCAATGGCCATCTG 344
|||||

DB 245 CTGGTGTGGCACTTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
DB 305 AGGATCACAAATGAGAAATTTTGTGGATGCTTACGAGAACTCCAACTCCACTGATTTGTA 364
QY 405 AGCTGGCGAGCGAGGTGAAGAGGCGCTGAAGTCTCTGTACAAATGAAGTCCCTGTCTG 464
DB 365 AGCTGGCGAGCAAGGTGAAGGACGCGCTGAAGTCTCTGTACAGCGGAGTCCCATCTCTG 424
QY 465 GGTCCCTACCAAGAAGTCTGCTGAACTGCTTCAGTTCAGGCGAGTGTCACTCGCTAC 524
DB 425 GGCCCTTACCAAGGAGTCTGCTGTGACGCGCTTTCAGGAGGCGAGCGTCACTGGCTAC 484
QY 525 TACTGTCTAGAGTTTCAGCATCCCTCCACACCTTGGCAGAAAGAGTTTGTATCGCGCATGGCT 584
DB 485 TACTGTCTGAGTTTCAGCATCCCGCAGCACCTTGTGTGAGGAGGCCGCGCTCATGGCC 544
QY 585 GTGGAGCGAGTTGTAAACATTTGCCACCCCGAGCACTGGGCACTGAATCTCTTCGTGTAACA 644
DB 545 GAGGAGCGGTGTATGCTATGCTGCCCCCGCGCGCTCTCCCTGAAGTCTTGTGTGTCACC 604
QY 645 TCTGTGTGGCTCTCCCTCATTTGACCCCAAGAACTGCTGCAGAGGACTCAGGACACAGCTGC 704
DB 605 TCAGTGTGGCTTTCCCGACGACTCCAAAACAGTACAGAGGACCCAGGACACAGCTGC 664
QY 705 AGTTTGGCTTCATGCCCATGGTCAGCAGTGACACGCTTCACTACCTTGGCTTCCCC 764
DB 665 AGCTTTGGCTTGCACGCGCGGTGTGAGTGTGCTGCTTCCACACGCGCGCTTCCCT 724
QY 765 ACAGTTCCTTACCGGGGATGCGCTGCGAGTGGTCTTGGGGGAGCGCGACTCT 824
DB 725 GACAGCGCTTACCGCTCATGCTGCGCGCTGCGAGTGGGCGCTGCGGGGAGCGCGACTCA 784
QY 825 GTCTGAGCTTCACTTCCGAGCTTTGATGTGCTGCTGCTGATGAGCATGCGAGTGC 884
DB 785 GTCTGAGCTTCACTTCCGAGCTTTGATGTGCTGCTGCTGAGCGCGCGCAGCGAC 844
QY 885 CTGGTCACTGATGATAGCTTACCGCTGAGCGCGCTTGGAAACCCCAAGCTGTGTGGCTGTGT 944
DB 845 CTGGTCACTGATGATAGCTTACCGCTGAGCGCGCTTGGAAACCCCAAGCTGTGTGGCTGTGT 904
QY 945 GGCACCTTCTACCTTCTACAGCTTACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
DB 905 GGCACCTTCT 964
QY 1005 ACCTGATTAACCAATATGACCGCGGACATCTCTGCTTTGAGGCGCACTTCTTCCAGCTG 1064
DB 965 ACCTGATTAACCAATATGACCGCGGCGCATCCCGCTTTGAGGCGCACTTCTTCCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGCTTTTGGAGTGACACCCCAAGGACATTTAGCAGCGCC 1124
DB 1025 CTTAGATGAGCAGCTGTGGAGCGCTTTACGTAAGCGCCAGGGGACATTTCAACAGCGCC 1084
QY 1125 TACTATCCAGGCGCACTACCGCGCCCAACATCACTGCACATGGAATCAAGTGGCGCCAC 1184
DB 1085 TACTATCCAGGCGCACTACCGCGCCCAACATCACTGCACATGGAATCAAGTGGCGCCAC 1144
QY 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGGTGGACCCCAACCTTACAGTG 1244
DB 1145 AACCGCATGTGAAGTGGAGTTCAAATCTTCTTCTACCTGTGGAGCGCGCTGCTGCG 1204
QY 1245 GGTCTCTGACCAAGACATGTGGAGATCAACGGGAGAAAGTACTGCGGTGAGAGTCC 1304
DB 1205 GGCACCTGCGCCCAAGGACTACGTTGGAGATCAATGGGGAGAAATACTGCGGAGAGGCTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTCATCTCG 1364
DB 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTCATCTCG 1324
QY 1365 TACACGGAACCGGGTCTCTAGCTAGTACCTCTCTTACGACTCCCAACGACCGCTGCCCA 1424
DB 1325 TACACGGAACCGGGTCTCTAGCTAGTACCTCTCTTACGACTCCCAACGACCGCTGCCCG 1384

1425 GGGATGTTTCATGTGCAAGACTGTGACGGTGCATCCGAAAGGAACTGCGCTGCGACGCTGG 1484
1435 GGGAGTTCACGTGCGCGACGGGGCGGTGTATCCGGAAGAGTGGCTGTGTATGCTGG 1444
1485 GCAGATGCCCGGATATAGTGTATGACGGTTACTGCGGATGCAATCCACCCACCACTGTC 1544
1445 GCCGATGTCACCGACACACAGCGATGAGTCAACTGCACTGCGACGCGGCCACCACTGTC 1504
1545 ACCTGCAAAACACAGTTCGACAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
1505 ACCTGCAAGAACAGTTCGACAGCCCTCTTCTGGGTCTGTGACAGTGTGAAAGCACTGC 1564
1605 GGGGACGGAAGTGACAGAGGGGCTGCGAGTGTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
1565 GGAGACAAACAGCGACGAGGGGTGAGTGTGTCGGGCCACAGACCTTCAGTGTTCGAAT 1624
1665 GGGAGTGTCTCCCTCAGAGCCAGAGAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCT 1724
1625 GGGAGTGTCTCTCGAAAGCCAGAGTGTCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
1725 GACGAGCTTCATGTGACAGCGTGAATGTGCTCTCTTGCAACCAATATACCTACCGCTGC 1784
1685 GACGAGCTTCCTGCCCAAGGTGAAGCTGTCTGCTTGTACCAAAACACCTACCGCTGC 1744
1785 CAAATGGGCTCTGTCTGACAAAGGGCAACCTCGAGTGTGATGGGAAGACGGACTGTAGC 1844
1745 CTCAATGGGCTCTGTGTGACAAAGGGCAACCTCGAGTGTGACGGGAAGGAGGACTGTAGC 1804
1845 GATGGCTCCGATGAGAAACCTGTGACTGTGGCTGGGATCCTTTACCAAAACAGGCTGCG 1904
1805 GACGGCTCAGATGAGAAGGACTCGACTGTGGGCTGGGTCATTTCACGACACAGGCTCGT 1864
1905 GTGGTGTGGTGGACGATCGGACGGGGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
1865 GTTGTGGGGACGAGATGGGATGAGGGGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1924
1965 CTGGGACAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
1925 CTGGGACAGGGCCACATCTGGGTGTCTCTCATCTCTCTCCAACTGGCTGTCTGCG 1984
2025 GCTCATGTCTTCAGGATGACAAATTTCAAGTACTCAGACTACAGATGTGGACGGCC 2084
1985 GCACATGTCTACATCATGATGACAGAGGATTCAGGTTACTCAGACCCCAACGCACTGGACGGCC 2044
2085 TTCTGGGTCTGTGGACACAGCAAGCGAGTGCCTCTGGGGTGCAGGAGTGAAGCTC 2144
2045 TTCTGGGTCTGACACACAGCCAGCGAGCGCCCTGGGGTGCAGGAGCGCAGGCTC 2104
2145 AAACGTATCATCACCAACCTTCTTCAATGATTTCACTTCGACTATGACATCGCCTTG 2204
2105 AAGCGCATCATCTCCACACCTTCTTCAATGACTTCACCTTCGACTATGACATCGCCTG 2164
2205 CTGAGCTGGAGAGTGGTGGAGTACAGACCGTGTGGCCGCCATCTGCTGCTGAT 2264
2165 CTGGAGCTGGAGAAACCGGAGAGTACAGCTCCATGGTGGGGCCCATCTGCTGCGGGAC 2224
2265 GCTACCATGTCTTCTGCTGGCAAGGCCATCTGGGTACAGGCTGGGGGCACAAAA 2324
2225 GCCTCCCATGTCTTCTGCTGGCAAGGCCATCTGGGTACAGGCTGGGGGCACACCCAG 2284
2325 GAGGAGGTACCGAGCGCTGATCTCTGCAAGAGGTGAGATCCGTGTTCATCAACAGACC 2384
2285 TATGGAGGCATCTGGCGGCTGATCTCTGCAAAAGGTTGAGATCCGCGTTCATCAACAGACC 2344
2385 ACCTGTGAGACCTCATGCGGACGAGATCACCCACGATGATGTGTGGGTTCCTC 2444
2345 ACCTGAGAGAACTCTCTGCGGACGAGATCACGCGCGCATGATGTGCGGTTCCTC 2404
2445 AGTGGGGGTGTGAGTCTCTGCGGGGTGACTCTGTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
2405 AGCGGCGGCTGGACTCTCTGCGAGGGTGTATCTCGGGGACCCCTGTCCAGCGTGGAGGG 2464

2505 GATGGCGGAATGTTCCAGGCTGCTGTGGTGAAGGCTGCGCTCAGAGGAAC 2564
2465 GATGGCGGATCTTCCAGGCCGCTGTGGTGAAGGCTGGGAGACGGCTGCGCTCAGAGGAAC 2524
2565 AAGCCAGGCGTGTACAAAGGCTCCCTGTAGTTTGGGACTGGATCAAAGAGCACTGGG 2624
2525 AAGCCAGGCGTGTACAAAGGCTCCCTGTAGTTTGGGACTGGATCAAAGAGCACTGGG 2584
2625 GTATAGCAGCATGGACAGACGCGACCAACCAACCCACAGGATGCCGACATGCACA 2684
2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACTCGGGGGCCACCCATCGTCCACC 2641
2685 CCTGGATACAGAGAGGAACACTGACGACATTTATGCTGTGGCTCTCCCGCCCAACACA 2744
2642 CCAGTGTGACG-CCTGCGAGGCTGGAGACTGGACCGTGACTGCACAGGCGCC-CCAGA 2699
2745 ACCGACTGTGAACCTGCAATCTTTAGGACTCAGAT 2780
2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

Search completed: February 21, 2004, 02:22:12
Job time : 812 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:07:32 ; Search time 179 seconds
(without alignments)
7658.860 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catggtgacggctgccggg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC

Gapop.10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883.2	60.6	3147	2	US-09-027-337-1
2	1883.2	60.6	3147	4	US-09-644-600-1
3	1883.2	60.6	3147	4	US-09-644-600-18
4	1530.8	49.3	2900	2	US-09-027-337-9
5	1530.8	49.3	2900	4	US-09-644-600-9
6	701.2	22.6	1553	4	US-09-280-116-10
7	306.8	9.9	434	4	US-09-702-705-1480
8	306.8	9.9	434	4	US-09-736-457-1480
9	173.2	5.6	796	4	US-09-280-116-107
10	113.6	3.7	2413	3	US-09-518-046-1
11	109.8	3.5	2544	3	US-09-518-046-3
12	101.6	3.3	1479	3	US-09-342-749-1
13	101.6	3.3	1479	4	US-09-691-840-1
14	101.2	3.3	1128	2	US-09-016-366A-20
15	101.2	3.3	1128	2	US-09-016-366A-15
16	100.6	3.2	1077	3	US-08-807-151-2
17	100.6	3.2	1077	4	US-08-807-151-2
18	99.6	3.2	1081	2	US-09-016-366A-22
19	99.6	3.2	1081	2	US-08-978-404B-17
20	98.4	3.2	2479	3	US-09-342-749-29
21	98.4	3.2	2479	4	US-09-691-840-29
22	98.2	3.2	735	3	US-09-079-970A-1
23	98.2	3.2	771	3	US-09-079-970A-4
24	98	3.2	1137	2	US-09-016-366A-18
25	98	3.2	1137	2	US-08-978-404B-13
26	94.2	3.0	2416	3	US-09-261-416-1
27	92.2	3.0	1605	2	US-09-000-446-1

28	91	2.9	980	4	US-09-023-942A-30	Sequence 30, Appli
29	91	2.9	1110	4	US-09-386-653A-1	Sequence 1, Appli
30	90.8	2.9	901	1	US-08-508-448C-9	Sequence 9, Appli
31	90.8	2.9	1460	4	US-09-370-838-80	Sequence 80, Appli
32	90.8	2.9	1462	4	US-09-370-838-55	Sequence 55, Appli
33	90.8	2.9	1517	1	US-08-508-448C-15	Sequence 15, Appli
34	90.8	2.9	2790	4	US-09-370-838-79	Sequence 79, Appli
35	90.4	2.9	1783	3	US-09-510-738A-188	Sequence 188, App
36	90.4	2.9	1783	4	US-09-861-966-188	Sequence 188, App
37	90.4	2.9	2363	4	US-09-742-703-3	Sequence 3, Appli
38	89.4	2.9	1154	2	US-09-016-366A-16	Sequence 16, Appli
39	89.4	2.9	1154	2	US-08-978-404B-11	Sequence 11, Appli
40	89.4	2.9	1212	4	US-09-620-312D-431	Sequence 431, App
41	88.8	2.9	1130	4	US-09-386-653A-8	Sequence 8, Appli
42	88.8	2.9	2038	3	US-09-008-271A-18	Sequence 18, Appli
43	88.8	2.9	2079	4	US-09-656-002-1	Sequence 1, Appli
44	88.6	2.9	1615	4	US-09-820-002-1	Sequence 1, Appli
45	88.4	2.8	696	1	US-08-508-448C-24	Sequence 24, Appli

ALIGNMENTS

RESULT 1
US-09-027-337-1
; Sequence 1, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 23..2589
; OTHER INFORMATION: cDNA sequence of TADG-15
US-09-027-337-1

Query Match	60.68;	Score 1883.2;	DB 2;	Length 3147;
Best Local Similarity	81.28;	Pred. No. 0;		
Matches 2223;	Conservative	0;	Mismatches 508;	Indels 5; Gaps 3;
Qy	45	GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGCTCTCAG	104	
Db	5	GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCGCGCAAGCGCGGAGGGGGCCGGAAG	64	
Qy	105	GACTTGGCGCGGAGCTCAAGTCAACTCCGGCTAGAGAACATGAATGGCTTTGAGGAG	164	
Db	65	GACTTGGCGCGGAGCTCAAGTCAACTCCGGCGCAGAGAAAGTGAATGGCTTTGAGGAA	124	
Qy	165	GGTGTGGAGTTCTGCTCGCAATGCGCAAGTGGAGAGCGAGGCGCCCGCCGCGC	224	
Db	125	GGCGTGGAGTTCTGCGAGTCAACACGTCAAGAGGTGGAAGCATGGCCCGGCGCGC	184	
Qy	225	TGGGTGGTGTGGTGGCAGTGTGTTCAGTTCCTCTTGTCTCTCCCTCATGTGGTGGCTTG	284	
Db	185	TGGGTGGTGTGGCAGCGCGTGTGATCGGCGCTCTCTTGTCTGTGGGATCGGCTTC	244	
Qy	285	CTGGTGGCATTCTCCATTATCGAATGTGCGGGTTCAAAAAGTCTTCAATGCCCATCTG	344	
Db	245	CTGGTGGCATTCTGAGTACCGGGAGTCCGCGTGTCCAGAAAGTCTTCAATGGCTACATG	304	
Qy	345	AGGATCACAAATCAGATCTTCTGGATGGGTATGAGACTCCACCTCCACAGAGTTTATC	404	
Db	305	AGGATCACAAATGAGAATTTTGTGGATGGCTACGAGAACTCCAACTCCACTGAGTTTGA	364	

405 AGCTGGCCAGCAGAGTGAAGAGGCGCTGAAGCTGCTGTACAAATGAAGTGCCTGTCTCTG 464
Db
365 AGCTGGCCAGCAGAGTGAAGAGCAGCGCTGAAGCTGCTGTACAGCGAGTCCCATTCCTG 424
QY
465 GGTCCCTTACACAAAGAGTGGCTGTAACTGCTTCAAGTGAAGGCAAGTGTATCGCTTAC 524
Db
425 GGCCTCTACCAAGAGAGTGGCTGTGAAGCGCTTTCAGCGAGGCGAGCTCATCGCTTAC 484
QY
525 TACTGTGTACAGAGTTCAGCATCCCTCCACACCTTGGCAGAGAGAGTTCATCGCCATGGCT 584
Db
485 TACTGTGTAGAGTTCAGCATCCCTCCAGAGACCTTGGTGAAGGCGCGCTCATGGCC 544
QY
585 GTGGAGCGAGTGTGAATATGGCCACCTTGGCAGAGCGGCACTGAATTCCTTGTGCTTAA 644
Db
545 GAGGAGCGGTAGTATGTCTGCTGCTGCGGCGCTTCCCTGAAGTCTTGTGGTCA 604
QY
645 TCTGTGGTGGCTTCCCTATGACCCAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 704
Db
605 TCAAGTGGTGGCTTCCCTACCGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGT 664
QY
705 AGTGTGGCTGTGATGCTGATGCTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 764
Db
665 AGTGTGGCTGTGATGCTGATGCTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 724
QY
765 AACAGTCCCTTACCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
Db
725 GACAGCGCTTACCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784
QY
825 GTGCTGAGCTTACCTTCCGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
Db
785 GTGCTGAGCTTACCTTCCGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
QY
885 CTGGTCACTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
Db
845 CTGGTCACTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 904
QY
945 GGCAGCTTCTGAGCTTCTGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
Db
905 GGCAGCTTCTGAGCTTCTGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 964
QY
1005 AGCTGATACCAATACTGACCGGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1064
Db
965 AACTGATACCAATACTGAGCGGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1024
QY
1065 CCAAGATGAGCAGCTGTGGCGCTTTTGAAGTGAACCCCAAGGAGATTTAGCAGCCCC 1124
Db
1025 CTTAGATGAGCAGCTGTGGAGCGCTTACGTAAGAGCCAGGAGATTTCAACAGCCCC 1084
QY
1125 TACTATCAGGCGCATACCGGCGCAATCACTGATGATGATGATGATGATGATGATGATGAT 1184
Db
1085 TACTATCAGGCGCATACCGGCGCAATCACTGATGATGATGATGATGATGATGATGATGAT 1144
QY
1185 AACCGGAACTGAGGTGGCTTCAAACTTCTATCTGCTGAGCCCAAGAGTACCAGTG 1244
Db
1145 AACCGAACTGAGGTGGCTTCAAACTTCTATCTGCTGAGCCCAAGAGTACCAGTG 1204
QY
1245 GGTCTCTGACCAAGGACTATGAGAGTCAAGCGGAGAGTACTGCGGTGAGAGTCC 1304
Db
1205 GGCACCTGCCCCAAGGACTAGTGGAGATCAATGGGAGAGATTAATGCGGAGAGTCC 1264
QY
1305 CAGTTTGTGGTGAAGCAACAGCAGCAGATTAAGTCACTTCACTTCACTTCACTTCACT 1364
Db
1265 CAGTTTGTGGTGAAGCAACAGCAGCAGATTAAGTCACTTCACTTCACTTCACTTCACT 1324
QY
1365 TACACGCAACCGGTTCTAGCTGAGTACCTTCTGCTGAGCTTCAACAGCAGCAGTCC 1424
Db
1325 TACACGCAACCGGTTCTAGCTGAGTACCTTCTGCTGAGCTTCAACAGCAGCAGTCC 1384
QY
1425 GGGATGTTTCAATGCAAGACTGGAAGGTCATCCGAAAGGAACTGCGCTGCGAGCGCTGG 1484
Db
1385 GGGCAGTTTCAATGCAAGACTGGAAGGTCATCCGAAAGGAACTGCGCTGCGAGCGCTGG 1444
QY
1485 GCAGACTGCCCCGATTAATAGTATGAGGTTACTGCGGATGCAATGCCACCCACAGTTC 1544

1445 GCCGACTGACCGACCAAGCGATGAGCTCAACTGAGTTCGACCGCCGCGCACAGATTTC 1504
QY
1545 ACGTCAAAAAACAGTTCGCAAGCCCTCTTCTGGTCTGTGACAGTGTCAACGACTGT 1604
Db
1505 ACGTCAAAAAACAGTTCGCAAGCCCTCTTCTGGTCTGTGACAGTGTGAACGACTGC 1564
QY
1605 GGGGACGGAAGTGAAGAGGCGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1664
Db
1565 GAGACAACAGCGAGCAGAGGCGTGCAGTGTGCGGCCACAGCTTCAGGTGTTCCTCAAT 1624
QY
1665 GGGAAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCT 1724
Db
1625 GGGAAAGTGTCTCTCGAAAGCCAGCAGTGCATGGGAAGGACGACTGTGGGAGCGGTTC 1684
QY
1725 GACGAGGCTTATGTGAAGCAGCTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1784
Db
1685 GACGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1744
QY
1785 CAAAATGGCTTGTGAGCAAGGCGACCTGAGTGTGATGGGAAGGACGAGTGTGCTGCT 1844
Db
1745 CTCATGGCTTGTGAGCAAGGCGACCTGAGTGTGAGCGGAAGGAGTGTGCTGCTGCT 1804
QY
1845 GATGCTTCGATGAGAAAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1904
Db
1805 GACGCTCAGATGAGAGGACTGCGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1864
QY
1905 GTGGTGTGGGCAAAATGCGGACGAGGCGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1964
Db
1865 GTGGTGTGGGCAAAATGCGGACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1924
QY
1965 CTGGGCGAGGCGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2024
Db
1925 CTGGGCGAGGCGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1984
QY
2025 GCTCATGCTTTCAGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGGAGCGGCC 2084
Db
1985 GCACACTGCTCATCGATGACAGAGGATTCAGGTACTCAGACCCACGAGTGGAGCGGCC 2044
QY
2085 TTCCTGGGTCTGCTGAGCAGAGCAGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2144
Db
2045 TTCCTGGGTCTGAGCAGCAGAGCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2104
QY
2145 AAACCTATCATCACCCACCTTCTCAATGATTTTCACTTCGACTATGATGCTGCTGCTG 2204
Db
2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTCGACTATGATGCTGCTGCT 2164
QY
2205 CTGGAGCTGGAGAGTGGTGGAGTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2264
Db
2165 CTGGAGCTGGAGAAACCGGCGAGTACAGCTCCATGGTGGCGCCCATCTGCTGCTGCTG 2224
QY
2265 GCTACCCATGCTTCTTCTGCTGCGCAAGGCCATCTGGGTACAGGCTGGGGGACACAAAA 2324
Db
2225 GCCTCCCATGCTTCTTCTGCTGCGCAAGGCCATCTGGGTACAGGCTGGGGGACACCCAG 2284
QY
2325 GAGGAGGTACCGGAGCGCTGATCTGCGAGAGGTGAGATCCGCTGCTCATCAACAGAGCC 2384
Db
2285 TATGAGGCACTGGCGCTGATCTGCAAAAGGCTGAGATCCGCTCATCAACAGAGCC 2344
QY
2385 ACCTGTGAGGACCTCATGCGCGAGCAGATTCACCCCAAGATGATGTGTGGGTTCCTC 2444
Db
2345 ACCTGTGAGAACTCTGCGCGAGCAGATTCAGCGCGCTGATGTGGGTGGGTTCCTC 2404
QY
2445 AGTGGGGTGTGGACTCTGCTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2504
Db
2405 AGCGGCGGTGGATCTCTGCGAGGCTGATTCGCGGGAGACCCCTGTCCAGGTGGAGGG 2464
QY
2505 GATGGCGGAATTTTCCAGGCTGGTGGTGGAGTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db
2465 GATGGCGGATCTTCCAGGCGGTGTGGTGGAGTGGGGGAGCGGCTGCGCTCAGAGGAAC 2524
QY
2565 AAGCCAGCGGTGTACACAGGCTTCTGTAGTTCGGGACTGATCAAGAGGACACTGGG 2624

Db 2525 AAGCCAGGCGGTATACAAAGGCTCCCTCTGTTTCGGAGCTGGATCAAAAGAGAAACACTGGG 2584
Qy 2625 GTATAGCAGCATGACAGACAGCGGACCAACAAACACCAAGGATGCCGACATGACACA 2684
Db 2585 GTATAGGCGCGGG--GCCACCAATATGTATACCTGCGGGGCCACCCATGCTCCACC 2641
Qy 2685 CTTGGATACAGGAGGAGAAACATGACGACATTTATGCTGTGGCTCCGCCCCCAACACA 2744
Db 2642 CCAAGTGTGACG-CCTGCAGCTGGAGACTGGACCGTGAAGTCAACAGCGGCC-CCAGA 2699
Qy 2745 ACCAGACTGTGAAGTCAATCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAAGTCAATCTCCAGGGCTCCAAAT 2735

RESULT 2
US-09-644-600-1
; Sequence 1, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hiroto
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-1

Query Match 60.6%; Score 1883.2; DB 4; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGCGCCGCAAGGCGGAGGGGGCCGAG 64
Qy 105 GACTTCGGCGGGGACTCAAGTCAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTCAACTCCCGGCAAGGAAAGTGAATGGCTTTGAGGAA 124
Qy 165 GGTGTGAGTTCTCGCTCGCAACAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 224
Db 125 GCGGTGAGTTCTCGCAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 184
Qy 225 TGGGT 284
Db 185 TGGGT 244
Qy 285 CTGT 344
Db 245 CTGT 304
Qy 345 AGGATCAAAATGAGATCTTCTGATGCGTATGAGAACTCCACTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGATGCGTATGAGAACTCCACTCCACAGAGTTTATC 364
Qy 405 AGCTTGGCCAGCGAGTGAAGAGAGCGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGT 464
Db 365 AGCTTGGCCAGCGAGTGAAGAGAGCGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGT 424
Qy 465 GGTCCCTACCAAGAGAGT 524

Db 425 GGCCCTTACCAAGAGTGGCTGTGAGCGCTTTCAGCGAGGCGCGTCAATCGCCTAC 484
Qy 525 TACTGTGAGATTGAGCATCCCGCCACACTGCGCAGAGAGTTGATGCGGCATGGCT 584
Db 485 TACTGTGCTGAGTTGAGCATCCCGCCACACTGCGCAGAGAGTTGATGCGGCATGGCT 544
Qy 585 GTGAGCGAGTTGTAACATTCGCCACCCGAGCAGCGGCACTGAAATCTTTCGTCTAAACA 644
Db 545 GAGAGCGCGTAGTCATGCTGCCCGCGGGCGCTCCCTGAAGTCTTGTGTGCTCACC 604
Qy 645 TCTGTGTGTGCGCTTCCCGCATTTGACCCCGAGAAATGCTGACAGAGACTCAGCAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTTCCCGCAGACTCCAAAACAGTACAGAGGACCCAGAGCAACAGCTGC 664
Qy 705 AGTTTTTGGCTGCATGCCCATGTTGAGCAGTACAGCGTTTCACTCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCTGCACGCCCGCGGTGTGAGCTGATGCGTTTCAACAGCGCGGGTTCCCT 724
Qy 765 AACAGTCCCTTACCCGCGCATGCGCTGCGAGTGGGTCTTGGGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCGCTGCGAGTGGGCCCTGCGGGGGAGCGCGACTCA 784
Qy 825 GTGCTGAGCTCACCCTTCCGAAAGTTTGTGATGCTGCTCCCTGTGATGAGCATGCGAGTGC 884
Db 785 GTGCTGAGCTCACCCTTCCGAGCTTTGACCTTGGCTCTTGGCAGCAGCGCGGAGCGAC 844
Qy 885 CTGCTCACCGTGTATGATAGCTGAGCCCATGGAACCCCGAGCGTGTGTGCGCTGTGT 944
Db 845 CTGCTGAGCTGTATCAACACCTTGAAGCCCATGAGCCCGTGTGTGAGTGTGTGTGT 904
Qy 945 GGCACCTTCTCACCTCTCTAACACTGACTTTCTCTCTCCAGAACGCTTCTTCTTGTG 1004
Db 905 GGCACCTTACCTCTCTCTAACACTGACTTTCTCTCTCCAGAACGCTTCTTCTTGTG 964
Qy 1005 AGCTGATTAACCAATGATGACCGCGGACATCTCTGCTTTTGGAGGCACTTCTTCTCAGCTG 1064
Db 965 ACATGTATAACCAACACTGAGCGCGGCATCCCGGCTTTGAGGCGCACTTCTTCTCAGCTG 1024
Qy 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCAAGGAGCATTTAGCAGCGCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGCGCTTACGTAAAGCCCGAGGAGCATTTCAACAGCGCC 1084
Qy 1125 TACTATCCAGGCGCACTACCGCCCAACATCAACTGCACTGCAATGGAATATCAAGGTGCCCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCGCCCAACATTTGACTGCACTGGAATATGAGGTGCCCCAAC 1144
Qy 1185 AACCGGAAGTGAAGTGGCTTCAAACTTCTTATCTGTGTGAGCCCAAGCTACCACTG 1244
Db 1145 AACCGCATGTGAAGTGAAGTTCAAATTTCTTCTGCTGGAGCGCGGCTGCTGGG 1204
Qy 1245 GGTCTCTGCAACAGGACTATGTGGAGTCAACGGGGAGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCGCCAGGACTGAGTGGAGTCAATGGGGAGAAATACTGCGGAGAGAGTCC 1264
Qy 1305 CAGTTGT 1364
Db 1265 CAGTTGT 1324
Qy 1365 TACACGAGACCGGGTTCTTAGCTGAGTACTCTCTTACAGCTCCAGAGCGCGTCCCA 1424
Db 1325 TACACGAGACCGGGTTCTTAGCTGAGTACTCTCTTACAGCTCCAGTACCGGCTCCCG 1384
Qy 1425 GGGATGTTCTATGTGCAAGACTGGAAGCTGATCCGAAAGGAAATGCGCTGCGAGCGGTGG 1484
Db 1385 GGGAGTTTCACTGCGCGCAGCGGGCGGTGTATCCGGAAGAGTGTGCTGTGTGTGTGT 1444
Qy 1485 GCAGATGCGCGGATTTATGATGAGCGTTTCTGCGAGTGAATGCAATGCCACCGAGTTC 1544
Db 1445 GCGGACTGCAACCGACAGCGATGAGCTCAACTGCACTGAGTTGCGAGCGCGGCGCACAGTTC 1504
Qy 1545 ACCTGCAAAACAGTTCTGAGAGCGCTTCTTGGGTCTGTGACAGTGTCAAGAGTCTGT 1604

Db 1505 ACCTGCAAGAACAAAGTTCTTCAAGCCCTCTCTCTGGCTCTGCACAGTGTGAACGACTGC 1564
QY 1605 GGGACGGAAGTACAGGAGGAGGCTGCAGCTGTCTCTCTGGAGTTTCAAGTGTTCACAT 1664
Db 1565 GGAGACACAGCAGCAGAGGGTGCAGTGTTCGGGCCAGACCTTCAGTGTTCACAT 1624
QY 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGACAACTGTGTGAGATGGGTCT 1724
Db 1625 GGGAAGTGTCTCGAAAGCCAGCAGTGTCAATGGGAAGACGACTGTGTGGGGAGCGGTCC 1684
QY 1725 GACGAGCTTCATGTGACAGCGTGTGAATGTCTCTCTGACCAAAATATATCTACCGCTGC 1784
Db 1685 GACGAGCTTCCTCCCAAGGTGAAGCTGTCTCTGACCAAAACACACTTACCGCTGC 1744
QY 1785 CAAAATGGCTCTCTCTGACGAGGGCAACCTCTGAGTGTGTGATGGGAAGACGAGTGTAGC 1844
Db 1745 CTCAATGGCTCTCTCTGACGAGGGCAACCTCTGAGTGTGTGACGGGAAGGAGTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGGTGGGTGCGATCTCTTTTACCAACAGGCTGCG 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGTGGGTGCGATCTCTTTTACCAACAGGCTGCG 1864
QY 1905 GTGGTGTGTCGACGATGCGGAGGAGGAGTGGCCCTGCGAGGTGAGCCTCCAGGCC 1964
Db 1865 GTTGTGGGGCAGCGATGCGGATGAGGGCAGTGGCCCTGCGAGGTGAGCCTGCGATGCT 1924
QY 1965 CTGGCCAGGGCCACTTGTGTGGGGCCCTCGCTCATCTCTCTCTGACTGGCTGTCTCTGCA 2024
Db 1925 CTGGCCAGGGCCACTTGTGGGTGTCTCTCATCTCTCTCTGACTGGCTGTCTCTGCC 1984
QY 2025 GCTCATGTCTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGAGCGGCC 2084
Db 1985 GCACACTGCTACATCGATGACGAGGATTCAGGTACTCAGACCCCAAGCGATGGAGCGGCC 2044
QY 2085 TTCTGTGGTCTGTGTGACACGAGCAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCTGTGGTCTGTGACGACGAGCAGCGAGCGCCCTCTGGGGTGCAGGAGCGCAGCTC 2104
QY 2145 AAAGTATCATACCCACCTCTCTCAATGATTCATCTCGATTCGACTATGATCGCGCTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTCTTCAATGATCTTCACTTCGACTATGATCGCGCTG 2164
QY 2205 CTGGAGCTGAGAGTGTGGTGGAGTACAGCACCGTCTGTGGCCCACTTCGCTGCTCTGAT 2264
Db 2165 CTGGAGCTGAGAAACCGGAGAGTACAGTCTCCATGTGTGGCGCCACTCTGCTGCGCGAC 2224
QY 2265 GCTACCCATGTCTTCTCTGTGCGAAGCGCATCTGGGTTCACAGGCTCGGGGCAACAAAA 2324
Db 2225 GCCTCCCATGTCTTCTCTGCGGCAAGGCCATCTGGGTACGGGCTGGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAAGGTGAGATCCGTGTATCAACACGAGCC 2384
Db 2285 TATGAGGCACTGGCGGCTGATCTCTGAAAGGTGAGATCCCGCTCATCAACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGCAGCATCACCCACGAAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAACCTCTCTGCGCAGCATCACCGCGCATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGGAATCTCTGCGAGGGTGAATCTCTGTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGTGGAATCTCTGCGAGGGTGAATCTCTGCGGGGACCCCTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGGATGTTCAGAGCTGTGTGGTGTGAGTGGGTGAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATGTTCAGAGCTGTGTGGTGTGAGTGGGTGAGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCAGCGGTGTACAGAGCTCTCTGTAGTCTGGGACTGGATCAAGGACACTCTGGG 2624
Db 2525 AAGCAGCGGTGTACAGAGCTCTCTGTAGTCTGGGACTGGATCAAGGAAACACTCTGGG 2584
QY 2625 GTATAGCAGCATGACAGACGCGACCAACCAACACAGGAGTGGCGGATGCCGATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACCTGTGGGGGCCACCCATCTCTCCACC 2641

QY 2685 CTGGATACAGGAGAGAACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACA 2744
Db 2642 CCAGTGTGCACG-CTTGAGGCTGGAGACTGGACCGCTGACTGCACAGCGCCC-CCAGA 2699
QY 2745 ACCGACTGTGAACTGCACTCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGCTCCAAAT 2735

RESULT 3

US-09-644-600-18/c

; Sequence 18, Application US/09644600

; Patent No. 6451500

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Tanimoto, Hirotooshi

; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease

; FILE REFERENCE: Overexpressed in Carcinomas

; CURRENT FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: US/09/644,600

; PRIOR FILING DATE: 1999-10-20

; PRIOR APPLICATION NUMBER: 09/027,337

; PRIOR FILING DATE: 1998-02-20

; NUMBER OF SEQ ID NOS: 98

; SEQ ID NO 18

; LENGTH: 3147

; TYPE: RNA

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Antisense of TADG-15

US-09-644-600-18

Query Match 60.6%; Score 1883.2; DB 4; Length 3147;

Best Local Similarity 81.2%; Pred. No. 0;

Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCGCGCAAGCCCGAGGGGCTCTCAG 104
Db 3143 GAGCGCCTCGGGGTACCATGGGGGCGGATCGGCCCGCAAGGGCGAGGGGCCCGAAG 3084
QY 105 GATTCGGCGCGGAGCTCAAGTACAACTCCCGCTAGAGAACTGAATGCTTTGAGGAG 164
Db 3083 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGCACAGAAAGTGAATGCTTTGGAGAA 3024
QY 165 GGTGTGAGTTCCTGCTGCGCAATCCCAAGAAAGTGAGAGCGAGGCCCGAGCGC 224
Db 3023 GCGTGTGAGTTCCTGCGAGTCAACAGTCAAGAGGTGAAAGATGATGGCCCGGGCGC 2964
QY 225 TGGGTGTGTGTGTGGCAGTGTCTTTCAGTTCCTCTTGTCTTCCCTCATGCTGGCTTG 284
Db 2963 TGGGTGTGTGTGTGGCAGCGGTGTGATCGGCCCTCTCTTGTGTGTGTGGGATCGGCTTC 2904
QY 285 CTGGTGTGGCACTTCCATTCATTCGGAATGCGGGTTCAAAAGTCTTCATGCGCATCTG 344
Db 2903 CTGGTGTGGCAATTCGAGTACCGGGACGTGCTGTCCAGAAAGTCTTCAATGCTCATG 2844
QY 345 AGGATCACAATCAGATCTTTCTGGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 2843 AGGATCACAATGAGAAATTTTGTGGATGCCCTACGAGAACTCCAACTCCACTGATTTGA 2784
QY 405 AGCTTGCCAGCAGGTGAAAGGAGCGCTGAAAGCTGCTGTACAAATGAAGTCCCTGCTG 464
Db 2783 AGCTTGCCAGCAAGGTGAAAGGAGCGCTGAAAGCTGCTGTACAGCGAGTCCCATTCCTG 2724
QY 465 GTTCCCTTACCAAGAGTGGCTGTAACTGCTTCACTGAGGCGAGTGTCACTGCGCTAC 524
Db 2723 GGGCCCTTACCAAGAGTGGCTGTGAGCGCTTTCAGCGAGGCGAGCGTCACTGCGCTAC 2664
QY 525 TACTGTGTGAGTTCAGCATCTCCCGCCACACTCGGCAAGAGGTGTGATCGCGCATGGCT 584

Db 2663 TACTGGTCTGAGTTACAGATCCCGCAGACCTGCTGGAGGAGCGCGTCAATGGCC 2604
QY 585 GTGGAGCGAGTTGTAACTTTGCCACCCCGAGCAGCGGACATGAATCCTTCGTGCTAAACA 644
Db 2603 GAGGAGCGCGTAGTCATGCTGCCCGCGCGCGCGCTCCCTGAAGTCTTTGTGGTCAACC 2544
QY 645 TCTGTGGTGGCTTCCCACTTGAACCCAGAAATGCTGCAGAGGACTCAGGAGCAACAGCTGC 704
Db 2543 TCAGTGTGGCTTTCCCAACGGACTCCAAACAGTACAGAGGACCCAGGACACAGCTGC 2484
QY 705 AGTTTTCCTCGATGCGCATGCTGAGTGTGAGTGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 2483 AGCTTTGGCTCTGACGCGCGCGGTGTGGAGCTGATGGGCTTTCACACGCGCGGCTCCCT 2424
QY 765 AACAGTCCCTACCCGCGCATGCCGCTGCCAGTGGTCTCGGGGGGAGCGCGACTCT 824
Db 2423 GACAGCCCCCTACCCCGCTCATGCCCCGTGCCAGTGGGCCCTTGGGGGGGACGCCGACTCA 2364
QY 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTCCTCCTGTGATGAGCATGGCAGTGAC 884
Db 2363 GTGCTGAGCCTCACCTTCCGAGCTTTGACCTTGGCTCCTGGACGAGCGCGCAGCGAC 2304
QY 885 CTGGTACCGTGTATGATAGCTGAGCCCCCATGGAACCCACGCTGTGTGGTGGCTGTGT 944
Db 2303 CTGGTGAACGCTGTACAAACCCCTGAGCCCCCATGAGCCCCACGCGCTGTGTGCTGTGT 2244
QY 945 GGCACCTTCTCACCTCCTCAACCTGACTTTCTCTCTCCCGAGACGCTTCTCCTGTGC 1004
Db 2243 GGCACCTTACCTCCTCCTCAACCTGACTTTCTCCTCTCCCGAGACGCTTCTGCTCATC 2184
QY 1005 ACGCTGATAACCAATACTGACCGCGCAGATCCTGGCTTTGAGGCACTTTCTTCCAGCTG 1064
Db 2183 ACATGATAACCAACTGAGCGCGGCATCCCGCTTTGAGGCCACCTTCTTCCAGCTG 2124
QY 1065 CCCAGATGAGCAGCTGTGGGGCTTTTGTAGTGACACCCAAAGGACATTTAGCAGCCCC 1124
Db 2123 CCTAGGATGAGCAGCTGTGGAGCGCTGTAGTAAAGCCCCAGGGGACATTTCAACAGCCCC 2064
QY 1125 TACTATCCAGCCACTACCCGCCCAACATCAACTGCACATGGAATCAAGGTGCCCCAAC 1184
Db 2063 TACTACCGAGCCACTACCCCAACATTTGACTGTGCACATGGAACATTTAGGTGGCCCCAAC 2004
QY 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGGTGGACCCCAACGCTACCAAGT 1244
Db 2003 AACCGCATGTGAAGTGTGAGCTTCAAAATCTTCTATCTGTGGAGCCCGCGCTGGG 1944
QY 1245 GCTCTCTGACCAAGGACTATGTGGAGTCAACCGGGAGAGTACTCGGTGAGAGTCC 1304
Db 1943 GGCACCTGCCCCCAAGGACTACGTGGAGTCAATGGGGAGAAATACTGCGGAGAGAGTCC 1884
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAAGATTAACAGTCCACTTCCATTTCTGATCACTCG 1364
Db 1883 CAGTTCTGTCTACCCAGCAACAGCAAGATTAACAGTTCGCTTCCACTCAGATCAGTCC 1824
QY 1365 TACAGGACACCGGGTTCCTAGTGAATGCTCTCTACGACTCCAAACGACCGGTGCCCA 1424
Db 1823 TACACCGACACCGGCTTCTTAGTGAATGCTCTCTCTACGACTCCAGTGAACCCATGCCCC 1764
QY 1425 GGGATGTTCTATGTGCAAGCTGAGCGTGTGATCCGAAAGGACTGCGCTGGAGCGGTGG 1484
Db 1763 GGGCAGTTTACGTGCCGACCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1704
QY 1485 GCAGACTGCCCGGATTAATGATGAGCGTGTACTTGGCGGATGCAATGCCACCCAGTTC 1544
Db 1703 GCCAGCTGCACCCAGCAGCGGATGAGCTCACTGAGTTGGAGCGCGGCGCACAGTTC 1644
QY 1545 ACGTGCAGAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGAGTGTCAACAGCTGT 1604
Db 1643 ACGTGCAGAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGAGTGTGAAACGACTGC 1584
QY 1605 GGGGACGGAAGTGAACGAGGAGGCTGAGCTGCTCTGCTGGAGTTCATAGTGTTCAT 1664
Db 1583 GGAACAAACGCGACGAGCGGGGTGAGTTGTCTCCGCCCGCAGACCTTTCAAGTGTTCAT 1524

QY 1665 GCGAAGTGTCTCCCTCAGACCCAGAAAGTGTAAATGGGAGAGCAACTGTGGAGATGGGTCT 1724
Db 1523 GCGAAGTGTCTCTCGAAAAGCGAGCTGCAATGGGAGAGCAGCTGTGGGAGCGGGTCC 1464
QY 1725 GAGAGGCTTCATGTGACAGCGTGAATGTCTCTTTGGACCAAAATATACCTACCGCTGC 1784
Db 1463 GACGAGGCTCTCTGCCCCAAGGTGAACGTCTCTCTTGTACAAAACACACTACCGTGC 1404
QY 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCTTGAGTGTGATGGGAGACGGACTGTAGC 1844
Db 1403 CTCATATGGGCTCTCTGTGAGCAAGGGCAACCTTGAGTGTGACGGAGAGGAGCTGTAGC 1344
QY 1845 GATGGCTCCGATCAGAAAACCTGTGCTGTGGCTGCGATCCTTTTCAACAAACAGGCTGC 1904
Db 1343 GACGGCTCAGATGAGAGGAGCTGCGACTGTGGCTGTGGCTCAATTCAGAGACAGGCTCGT 1284
QY 1905 GTGTTGGTGGCAGCAATGCGACGAGGGGAGTGGCCCTGGCAGGTGAGCTCCACGCCC 1964
Db 1283 GTTGTGGGGCACGATGCGGATGAGGGGAGTGGCCCTGGCAGGTAAAGCTGCATGCT 1224
QY 1965 CTGGGCCAGGGCCACTTGTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGTGCTGCA 2024
Db 1223 CTGGGCCAGGGCCACATCTGGGTGTCTCTCATCTCTCCCACTGGCTGTCTGCTGCC 1164
QY 2025 GCTCATTTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1163 GCACACTGCTTACATCGATGACAGAGGANTCAGGTACTCAGACCCCGCAGAGTGGACGGCC 1104
QY 2085 TTCTGGGTCTGTGGACCAAGACGAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAAGCTC 2144
Db 1103 TTCTGGGGTTGCACGACCAAGACGAGCGCGCCCCCTGGGGTGCAGGAGCGCAGGCTC 1044
QY 2145 AACGATATCATACCCACCCCTTCTTCAATGATTTTCACTTGAATGATGATGCGCTTG 2204
Db 1043 AACGCACTATCTTCCACCCCTTCTTCAATGACTTCACTTGGACTATGATCATCGGCTG 984
QY 2205 CTGAGCTGGAGAAAGTCTGGTGGAGTACAGCACCGTCTGTGGCGCCCATCTCCCTGGCTGAT 2264
Db 983 CTGAGCTGGAGAAACCGGACAGAGTACAGCTCATGTGTGGCGGCCCATCTGCTTGGCGGAC 924
QY 2265 GCTVACCCATGTCTTCCCTGTGTGCAAGGGCCATCTGGGTCAACGGCTGGGGGCAACAAAA 2324
Db 923 GCTCTCCATGTCTTCTTCCCTGCGGCAAGGCCATCTGGGTCAACGGCTGGGGACACACCCAG 864
QY 2325 GAGGGAGGTACCGAGGCGTGTGCTGCGAAGAGGTGAGATCCGTGTCTCATCAACAGACCC 2384
Db 863 TATGGAGGCACTGGCGGCTGTATCTTGCAAAAGGGTGAATCCGCGCTCATCAACAGACCC 804
QY 2385 ACCTGTGAGGACCTCATCGCGACAGATCACCCCGACGATGATGTGTGGGTTCCTC 2444
Db 803 ACCTGCGAGNACCTCTGCGCAGCAGATCACGCGCGCATGATGTGCTGGGCTTCTC 744
QY 2445 AGTGGGGGTGTGACTCTCTGCGCAGGGTGAATCTGTGTGGCCCCCTTTGTCAAGCGCGGAGAAA 2504
Db 743 AGCGGGCGGTGACTCTCTGCGCAGGGTGAATCTGCGGGGACCCCTGTCCAGCGTGGAGGCG 684
QY 2505 GATGGCGGATGTTCCAGGCTGTGTGTGAGCTGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 683 GATGGCGGATGTTCCAGGCGGTGTGGTGAAGCTGGGAGACGGCTGCGCTCAGAGGAAC 624
QY 2565 AAGCCAGGGCTGTACAAAGGCTCCTGTAGTTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 623 AAGCCAGGGCTGTACACAGGCTCCTCTGTTTCGGGACTGGATCAAGAGAACACTGGG 564
QY 2625 GTATAGCAGCATGGACAGACAGCCGACCAAAACACCCACAGGGATGCCCGACATGCACA 2684
Db 563 GTATAGGGCGGGG---GCCACCAATGTGTACACTGTGGGGCCACCCATCGTCCACC 507
QY 2685 CTTGGATACAGGAGAGGAACACTGACATTTATGCTGTGGCTCCCCCCCCCAACACA 2744
Db 506 CCAGTGTGACG---CTTGCAGGCTGGAGACTGGACCGCTGACTGCAACAGCGCCC---CCAGA 449

QY 2745 ACCGAGACTGTGAACATGATCTTCTAGGACTCAGAGT 2780
|||
Db 448 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 413

RESULT 4

US-09-027-337-9
; Sequence 9, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 9
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SNC19 mRNA sequence (U20428)
US-09-027-337-9

Query Match 49.3%; Score 1530.8; DB 2; Length 2900;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2011; Conservative 0; Mismatches 502; Indels 29; Gaps 15;

QY 222 CGCTGGTGTGTGGGAGTGTCTTTCAGCTTCCTCTTGTCTTCCCTCATGGTGGC 281
Db 1 CGCTGGTGTGTGGGAGTGTCTTTCAGCTTCCTCTTGTCTTCCCTCATGGGATCGC 60
QY 282 TTGCTGTGTGGGACATTCATATCGAATGTGGGCTTCAAAAGTCTTCAATGGCCAT 341
Db 61 TTCTGTGTGGGACATTCATATCGAATGTGGGCTTCAAAAGTCTTCAATGGGCTAC 120
QY 342 CTGAGGATCAAAATGAGATCTTTCTGATGCGTATGAGAACTCCACTCCACAGAGTTT 401
Db 121 ATGAGGATCAAAATGAGATTTTGTGATGCTTACGAACTCCAACTCCACTGAGTTT 180
QY 402 ATCAGCTGGCCAGCCAGGTGAAGGCGCTGAAGCTGTGTGTAATGAATGCCCTGTTC 461
Db 181 GTAAGCTGGCCAGCCAGGTGAAGGCGCTGAAGCTGTGTGTAATGAATGCCCTGTTC 240
QY 462 CTGGTCCCTACACAGAGTGTGGTGTAACTGCTTCACTGAGGAGTGTCTATGCC 521
Db 241 CTGGGCCCCCTACCAAGGAGTGGCTGTGACGGCTTTCAGCGAGGCGAGCTCATGCC 300
QY 522 TACTACTGGTCAGAGTTTCAAGATCCCCCAACCTGGCAGAGAGTTGATCGCGCCATG 581
Db 301 TACTACTGGTCAGAGTTTCAAGATCCCCCAACCTGGTGTGAGGAGGCGAGCGGTATG 360
QY 582 GCTGTGAGCGAGTTGTAACTTGGCAGCCCGAGCAGCGGCACTGAAATCTTCTGTGCTA 641
Db 361 GC-CAGGAGCGGTAGTCACTGCTCCCGCGGCGGCGCTCCCTGAAATCTTGTGTGTC 419
QY 642 ACATCTGTGTGGCTTCCCATTTGACCCAGATGCTGAGAGGACTCAGGACAAACAGC 701
Db 420 ACCTAGTGTGTGGCTTCCCATTTGACCCAGATGCTGAGAGGACTCAGGACAAACAGC 479
QY 702 TGCAGTTTGGCTTGCATGCCCATGTTGTCAGTGTGACAGCTTCACTTACCTTGGCTTC 761
Db 480 TGCAGTTTGGCTTGCATGCCCATGTTGTCAGTGTGACAGCTTCACTTACCTTGGCTTC 537
QY 762 CCCAAGCTCCCTACCCGGGCGATGCCCGTGTGCGAGTGGTCTGCGGGGGAGCGCCGAC 821
Db 538 CTTGACAGCCCTTACCCCGCTCATGCCCGCTGCGAGTGG-----GCTGCGGGGACGCGAC 592
QY 822 TCTGTGCTGAGCCCTACCTTCCGAGCTTTGATGTGCTTCTCTGTGATGATGAGCTC 881
Db 593 GCACTGTGAGCTACTCGAGCTGACTCGCAGCTTGACTGTGCGCTTCGAGCGCGGCGAGC 652

QY 882 GACCTGGTCACCGTGTATGATAGCTGAGCCCATGGAACCCCAACGCTGTGGTGGGCTG 941
Db 653 GACCTGGTGA-CGTGTACAACACCCCTGAGCCCCATGGAGCCCCACGC---CTGGTGAAGT 708
QY 942 TGTGGCACCTTCTTACCCCTCCTACAACTGACTTTTCTCTCC--TCCAGAAAGTCTTCC 999
Db 709 TGTGGCACCTTACCTCCTCCTTACAACTGACTTCCCTCCCTCCCAAGAGTCTCTGC 768
QY 1000 TTGTACGCTGATAACCAATACTGACCGGGGACATCTTGGCTTTGAGGCCACTTTCTTC 1059
Db 769 TCATCACACTGTAAACCAACACTGAC--GCGGCATCCCGCTTTGAGGCCACTTTCTTC 826
QY 1060 AGCTGCCCAAGATGACAGCTGTGGCGGCTTTTGTAGTGCACCCCAAGGAGACATTTAGCA 1119
Db 827 AGCTGCTTAGATGACAGCTGTGGAGGCGCTTAGTAAAGCCAGGAGACATTTACACA 886
QY 1120 GCCCTTACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGAAATPATCAAGTGC 1179
Db 887 GCCCTTACTATCCAGGCCACTACCCCAACCAATGACTGACATGCAATGGAATTTAGGTGC 946
QY 1180 CCAACACCGGACGTAAGGTGGCTTCAAACTCTTCTATCTGTGGTGGACCCCAAGTAC 1239
Db 947 CCAACACCGAGCATGTGAAGGTGGCTTCAAACTCTTCTATCTGTGGTGGACCCCGGTGC 1006
QY 1240 CAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAACTACTGCGGTGAGA 1299
Db 1007 CTGCGGGCACTGCCCCAAGGACTAGTGGAGATCAATGGGAGAACTACTGCGGAGAGA 1066
QY 1300 GGTCCCAAGTTTGTGTGAGAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATC 1359
Db 1067 GGTCCCAAGTTTGTGTGAGAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATC 1126
QY 1360 ACTGTGTACAGGACACGGGTTCTTAGCTGAGTACCTCTCTACGACTCCCAAGACCCGT 1419
Db 1127 AGTCTTACCCGACACCGGCTTTTAGCTGAATACCTCTCTACGACTCCAGTGAACCAT 1186
QY 1420 GCCCAGGAGTTTCTATGTCAAGACTGCGAGCTGATCCGAAAGAACTGCGCTGCGAGC 1479
Db 1187 GCCCGGGGCGATTCAGTGGCGCACCGGGCGGTATCCGGAAGGAGCTGCGCTGTGATG 1246
QY 1480 GCTGGGAGACTCCCGGATTAAGTGTAGAGGCTTACTGCCGATGAAATGCCACCCACC 1539
Db 1247 GCTGGGG-GACTGACCCGACACAGCGATGAGTCAACTGCACTGCGACCGCGGCCACC 1305
QY 1540 AGTTTACGTCAAAAACCACTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACG 1599
Db 1306 AGTTTACGTCAAGAGCAAGTTCTGCAAG---CTCTTCTGGGTCTGCGACAGTGTGACG 1362
QY 1600 ACTGTGGGAGCGGAAGTGAACGAGGCGGTGCAAGTGTCTCTG-CGCGAGTTTCAAGTGT 1658
Db 1363 AGTGGGAGACAACAGCAGCAGCAGGGTTGCAATTTGTCGGGACCCAGACCTTTCAGGTGT 1422
QY 1659 TCCAATGGGAGTGTCTCCCTCAGAGCCAGAAAGTGAATGGGAGGACAACTGTGGAGAT 1718
Db 1423 TCCAATGGGAGTGTCTCCCTCAGAGCCAGAAAGTGAATGGGAGGACAACTGTGGGAC 1482
QY 1719 GGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTGCTCTCTTGCACCAATATACCTAC 1778
Db 1483 GGGTCTGACGAGGCTTCCTGCCCAAGGTGAAGTGTGCTACTTGTACCAACACACCTAC 1542
QY 1779 CGTGTCCAAATGGCTCTGTCTGAGCAAGGCAACCTCTGAGTGTGATGGGAAGACGGAC 1838
Db 1543 CGTGTCTCAATGGGCTCTGCTTTGAGCAAGGCAACCTCTGAGTGTGACGGGAAGGAGAC 1602
QY 1839 TGTAGGATGGCTCCGATGAGAAAACTGTGACTGTGGGCTGCGATCTTCTTACCAACAG 1898
Db 1603 TGTAGGATGGCTCCGATGAGAAAACTGTGACTGTGGGCTGCGATCTTCTTACCAACAG 1662
QY 1899 GCTCGCTGTGTGGTGGCAGAACTGCGAGCGAGGAGTGGCCCTGCGGAGGTGAGCTC 1958
Db 1663 GCTCGCTGTGTGGTGGCAGAACTGCGAGCGAGGAGTGGCCCTGCGGAGGTGAGCTC 1722

QY 1120 GCGCTACTATCCAGGCGCACTACCGCCCAACATCAATGACATGAATATCAAGTGC 1179
Db 887 GCGCTACTATCCAGGCGCACTACCGCCCAACATGAGTGCACATGAATATCAAGTGC 946
QY 1180 CCAACACCGGACCGTGAAGTGGCTTCAACATCTTCTATCTGCTGAGACCCCAACGATC 1239
Db 947 CCAACACCGGACCGTGAAGTGGCTTCAACATCTTCTATCTGCTGAGACCCCGCTGC 1006
QY 1240 CAGTGGGCTCTGACCAAGGACTATGAGAGATCAACGGGAGAACTACTGCGGTGAGA 1299
Db 1007 CTGCGGCGACCTGCGCCCAAGGACTACGTGGAGATCAATGGGAGAACTACTGCGGAGAGA 1066
QY 1300 GGTCCCACTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTTCTGATC 1359
Db 1067 GGTCCCACTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTTCTGATC 1126
QY 1360 ACTGTCACACGACACCGGGTCTTAGTCTGAGTACCTCTCTAGACTTCCAAACGACCGGT 1419
Db 1127 AGTCCCTACACGACACCGGGTCTTAGTCTGAGTACCTCTCTAGACTTCCAAACGACCGGT 1186
QY 1420 GCGCAGGAGTGTATGTGCAAGACTGACCGGTGATCCGAAAGAACTGCGCTGCGAGC 1479
Db 1187 GCGCGGGGAGTGTACGTGCGCACCGGGCGGTGTATCCGGAAGAGCTGCGCTGTGATG 1246
QY 1480 GCTGGGCGAGCTGCCCGGATTATAGTATGAGCTTACTGCGGATGCAATGCCACCCACC 1539
Db 1247 GCTGGGC-GACTTGACCGACACAGGATGAGTCAACTGCACTTGCAGCGCGGCCACC 1305
QY 1540 AGTTCACGTGCAAAAACAGTTCTTGCAGGCCCTCTTCTGGGTCTGTGACAGTGTCAAG 1599
Db 1306 AGTTCACGTGCAAGAGCAAGTTCTGCAAG--CTTCTTGGGTCTGCGACAGTGTGAAAG 1362
QY 1600 ACTGTGGGACCGAAGTACGAGGAGGCTGACGTGCTCTG--CTGGGAGTTTCAAGTGT 1658
Db 1363 AGTGTGGGACCAACGCGACGAGCGGTGTGATTTGTCCGACCCAGACCTTCAAGTGT 1422
QY 1659 TCCATATGGGAAGTGTCTCCCTCAGAGCCAGAAAGTGTATGGGAAGCAAACTGTGAGAT 1718
Db 1423 TCCATATGGGAAGTGTCTCTGAAAGCCAGCAGTGTCAATGGGAAGCAAGTGTGGGAC 1482
QY 1719 GGTCTGTACAGAGCTTCAATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTAC 1778
Db 1483 GGTCTGTACAGAGCTTCTGCGCCCAAGTGAACGTCTCACTTGTACCAAAACACCTAC 1542
QY 1779 CGTGTCAAAATGGCTCTGTGACAGGGCAACCTGAGTGTGATGGGAAGACGAGAC 1838
Db 1543 CGTGTCTCAATGGGTCTGTGACAGGGCAACCTGAGTGTGACGGGAAGGAGAC 1602
QY 1839 TGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGTGCGATCCCTTTACCAACAG 1898
Db 1603 TGTAGCGACGCTCAGTGAAGAGTGTGCGACTGTGGGTGCGGTCAATTCAGAGACAG 1662
QY 1899 GCTCGGTGTTGGTGACAGATGGGACGAGGGGAGTGGCCCTGCGAGGTGAGCCCTC 1958
Db 1663 GCTCGGTGTTGGTGACAGATGGGATGCGATGTGGGTGCGGTCAATTCAGAGACAG 1722
QY 1959 CAGCCCTTGGCGAGGCGCACTTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGGTC 2018
Db 1723 CATGCTTGGGCGAGGCGCACTGCGGTGCTTCCCTCATCTCTCCAACTGGCTGGTC 1782
QY 2019 TCTGAGCTCATTTCTTCAAGATGACAAAAATTCMAAGTACTCAGACTACAGATGTGG 2078
Db 1783 TCTGCGCGACACTGCTACATCATGATGACAGAGGATTCAGGTACTCAGACCCCAAGC -CAGG 1840
QY 2079 ACGGCTTCTGGGTCTGTGACAGAGCAGAGCGAGTGGCTCTGGGTGCGAGGCTG 2138
Db 1841 ACGGCTTCTGGGTCTGTGACAGAGCAGAGCGAGGCA--GGCCCTGGGTGCGAGGCGC 1898
QY 2139 AAGCTCAACGATCATCACCCCTTCTTCAATGATTTCACTTCTGACTATGACATC 2198
Db 1899 AAGCTCAACGCGATCATCTCCACCCCTTCTTCAATGACTTCACTTCTGACTATGACATC 1958
QY 2199 GCCTTGTGGAGTGGGAAGTGGGTGAGTACAGCACCGTGTGCGGCCCATCTGCGCTG 2258

Db 1959 GCCTTGTGGAGTGGAGAAACCGGACAGATACAGCTTCAATGTGGGCCCATCTGCTG 2018
QY 2259 CTTGATGTACCCATGCTTCTCCCTGTGCAAGGCCATCTGGGTCAAGCTGGGGGAC 2318
Db 2019 CCGACGCTCCCATGCTTCTCCCTGCGGCAAGGCCATCTGGGTCAAGCTGGGGAC 2078
QY 2319 ACAAAGAGGAGGTACCGGAGCGCTGATCTCTGAGAGGGTGAATCCGTGTCTCAAC 2378
Db 2079 ACCAGTATGAGGCACTGGCGGCTGATCTCTCAAAAGGGTGAATCCGCTCATCAAC 2138
QY 2379 CAGACCACTGTGAGGACCTCATGCGCAGACATCAACCCACGAAATGATGTGTGGGT 2438
Db 2139 CAGACCACTGTGAGGACCTCATGCGCAGACATCAACCCACGAAATGATGTGTGGGT 2198
QY 2439 TCTCTAGTGGGGTGTGAGTCTCTGCGCAGGGTGAATCTCTGTGGCCCTTGTCAAGCGC 2498
Db 2199 TCTCTAGCAGCGCGGTGAGTCTCTGCGCAGGGTGAATCCGCGGACCCCTGTCCAGCGT 2258
QY 2499 GAGAAAGATGGCGAATGTTCCAGCTGTGTGAGTGGGTGAAGGTGGCTGCTCAG 2558
Db 2259 GAGCGGATGGCGGATCTTCCAGCGCGTGTGAGTGGGTGGGAG--ACGCTGCGCTCAG 2317
QY 2559 AGGAACAGCCAGCGGTGTACAAAGCTCCCTGTAGTTTGGGACTGGATCAAAAGACAC 2618
Db 2318 AGGAACAGCCAGCGGTGTACAAAGCTCCCTGTGTTTGGGATGGATCAAAAGAGAC 2377
QY 2619 ACTGGGTATAGCAGCATGGACAGACCCACCAAAACCCACAGGGATGCCGACA 2678
Db 2378 ACTGGGTATAGGGCGCGG--GCCACCAATGTGTACACCTGCGGGGACCCATCG 2434
QY 2679 TGCACCTGATACAGGAGAGACACTGACGACATTTATGTGTGCGCTTCCCGCCCC 2738
Db 2435 TCCACCCAGTGTGCACGCTGCGAGGTGGAGACTGCGGACCGGTGACCTGCACGCGC 2494
QY 2739 AACCAACCCAGACTGTCAACT 2760
Db 2495 CCGAACAATACACTGTGACT 2516

RESULT 6

US-09-280-116-10
; Sequence 10, Application 'US/09280116A'
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1553)
; OTHER INFORMATION: n = a, t, c, or g
US-09-280-116-10

Query Match 22.6%; Score 701.2; DB 4; Length 1553;
Best Local Similarity 71.1%; Pred. No. 1.2e-177;
Matches 1092; Conservative 0; Mismatches 351; Indels 92; Gaps 9;
QY 1659 TCCAAATGGGAAGTGTCTCCCTCAGAGCCAGCAAGTGTAAATGGGAAGGACAACTGTGGAGAT 1718
Db 1 TCCGATGGAAGTGCCTTTCGAAAAGCCAGCAGTGTCAATGGGAAGGACGACTGTGGGAC 60
QY 1719 GGGTCTGACGAGGCTTTCATGTGACGCGTGAATGTCTCTTGTGCAACCAATATACCTAC 1778

Db	1136	CTCGCTTCCTCAGCTCCAAAGTGGAGCTGGGAGGTAGAAAGGGAGACACTGGTGGTTC	1195
Qy	2829	TAGCGGCCAGCCTGGGGGCAAGGGTTTGATGCGAGCGCTTCCCGCTCTAGCCCTCAGACTG	2888
Db	1196	TACTGACCCAACTGGGGGCAAGGGTTTGAAGACACAGGCTTCCCGCGCAGCCCAAGCTG	1255
Qy	2889	GGTGAAGATGATGCTGTCCGGAGAGCTGCTTC-----CAACTGTCTATTGAG	2935
Db	1256	GGCGAGGCGGTTTGTGCATATCTGCCTCCCTGCTCTAAGGAGCAGCGGAACGGAG	1315
Qy	2936	CTCCCGGAGCCCTA-----TGGGAGGAGGGGCTCAGGG	2969
Db	1316	CTTGGGGCGCTCTCAGTGAAGGTGTGGGGCTGCCGGAATCTGGGCTGTGGGGCTTTGGG	1375
Qy	2970	TCACCTCTTTTCAGAGAGCGCCAGCCCTAGGAACCCAGAAAAAGAGTGGTACCTAAGGCTG	3029
Db	1376	CCAGGCTCTTGAGAGGCCCAGGCTCGGAGGACCTTGGAAAAACAGACGGGTCTGAGACTG	1435
Qy	3030	AAATGTTTTGCTGTTGCCAGGGGTGG-----GTATTGTGAGAGTAAA	3071
Db	1436	AAATGTTTTTACCAGCTCCCGAGGTGCACCTTCAGTGTGTGTAATTTGTGTAATGAGTAAA	1495
Qy	3072	ACATTTTATTTCTTTTAAAAAATAAAAAAAAAAAAA	3106
Db	1496	ACATTTTATTTCTTTTAAAAAATAAAAAAAAAAAAA	1530

RESULT 7
US-09-702-705-1480
; Sequence 1480, Application US/09702705

Query Match	9.9%;	Score 306.8;	DB 4;	Length 434;
Best Local Similarity	82.1%;	Pred. No. 2e-72;		
Matches 353;	Conservative 0;	Mismatches 77;	Indels 0;	Gaps 0;
Qy	1083	GGCGGCTTTTGTAGTGACACCCAAAGGACATTATAGCAGCGCCCTACTACTATCCAGGCCACTAC	1142	
Db	1	GGAGGCCGCTTACGTAAGGCCCAAGGGGACATTCAACAGCGCCCTACTATCCAGGCCACTAC	60	
Qy	1143	CCGCCCAACATCAACTGTCACATGGAAATATCAAGGTGCCCAACAAACCGGAACGTGAAGGTG	1202	
Db	61	CCACCCAAACATTGACTGTCACATGGAAACATTGAGGTGCCCAACAAACCAAGCAGTGTGAAGGTG	120	
Qy	1203	CGCTTCAAACTCTTTCTATCTGTGTGACCCCAACGTCACAGTGGGCTCCTGCAACCAAGGAC	1262	
Db	121	CGCTTCANAATTCCTTACTCTGTGTGAGCCGGCGTGCTGCGGCGACCTGCCCAAGGAC	180	
Qy	1263	TATGTGAGATCAACCGGGAGAGTACTCGGTGAGAGGTCCAGTTTGTGTGAGCAGC	1322	
Db	181	TACGTGGAGATCAATGGGAGAAATACTCGAGAGAGAGGTCCAGTTTGTGTGAGCAGC	240	

Query Match	9.9%;	Score 306.8;	DB 4;	Length 434;
Best Local Similarity	82.1%;	Pred. No. 2e-72;		

QY 2284 CTGGCAAGGCCATCTGGGTCAAGGCTGGGGGCAACAAGAGAGGAGGTACCGGAGCG 2343
DB 625 CCGGCTGCACTGCTGGATTAGGGCTGGGGCGCTTGGCGAGGGCGGCCCATCAGCA 684
QY 2344 TGATCTCTCAGAAAGGTGAGATCGGTGATCAACACAGACACCTGTGAGGACCTCATGC 2403
DB 685 AGCTCTCTCAGAAAGTGGATGTGCANTTGTATCCACAGGACCTGTGCAGCGAGGTCTATC 744
QY 2404 CGCAGCAGATCACCCACCAAGATGAT 2428
DB 745 GCTACCAAGTACGCCACGCAATGCT 769

RESULT 10
US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1

Query Match 3.7%; Score 113.6; DB 3; Length 2413;
Best Local Similarity 52.6%; Pred. No. 2.1e-20;
Matches 388; Conservative 0; Mismatches 314; Indels 36; Gaps 5;

QY 1900 CTCGGTGGTGTGGACGAAATGGCGACGAGGGAGTGGCCCTGGAGGTGAGCTTC 1959
DB 790 CACGCATGCTGGGTGGAAACATGCTTGTCTCGCAGTGGCCCTGGCAGGCGACGCTTC 849
QY 1960 ACGCCCTGGGGCAGGGCCACTTGTGGGGCTCGCTCATCTCTCTGACTGGCTGTCT 2019
DB 850 AGTTCAGGGCTA---CCACTGTGGGGGGCTGTGTATCAGCCCTCTGGATCATCA 906
QY 2020 CTGAGCTCAATGTCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGA 2079
DB 907 CTGCTGCACACTGTGTTTATGACTGTGTACCTCCCAAGTCAATGACCATCCAGTGGTC 966
QY 2080 CGGCTTCTGGGTCTGTGACACAGCAGCAGCGAGTGTCTTGGGGTGCAGAGCTGA 2139
DB 967 TAGTTTCCCTG-----TTGGCAATTCACGCCCATCCCACTTGTGGAGAAG----- 1013
QY 2140 AGCTCAAAAGCTATCATCACCCACCTCTTCAATGATTTTCACTTCGACTATCAGATCG 2199
DB 1014 -----ATTGTCTACACAGCAAGTACAGCCAAAGAGCTGGGCATGACATCG 1062
QY 2200 CTTGTGGAGCTGGAGAGTCTGGTGGAGTACAGACCGTCTGGCCCTCATCTGCTGC 2259
DB 1063 CCTTATGAGCTGGCCGGGCACTCAGTTCAATGAAATGATCCAGCTGTGTGCTGC 1122
QY 2260 CTGATGCTACCCATGCTTCCCTCTGTCAGAGGCACTTGGGTACAGGCTGGGGGACA 2319
DB 1123 CCAACTCTGAAGAGAACTTCCCGATGGAAGTGTGCTGGACGTGAGGATGGGGGCCA 1182
QY 2320 CAAAAGAGGAGGTACCGGAGCGTGTCTGTCAGAGAGGGTGTGAGATCGGTCTCATCAAC 2379
DB 1183 CAGAGGATGGAGGTGACGCTTCCCTCTGCTTGAACACCGCGGCCGTCCCTTTGATTCCA 1242

QY 2380 AGACCACCTGTGA-----GGACCTCATGCCGACGACATCACCCACGAATGATGTG 2433
DB 1243 ACAAGATCTGCAACACACAGGAGCGGTACGGTGGGATCATCTCCCCCTCCATGCTCTCG 1302
QY 2434 TGGTTTCTCAGTGGGGGTGGGACTCTCCAGGGTGAAGTCTGTGGTGGCCCCCTTGTCAA 2493
DB 1303 CGGGCTACTCAGCGTGGCGTGAACAGTCCAGGGGGAAGCGGGGGCCCCCTGTGT 1362
QY 2494 GCGGGGAGAAAGATGGCGGAATTTCCAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2553
DB 1363 GTCAGAGAGAGGAGGCTGTGGAAGTT--AGTGGAGAGCGACAGCTTTGGCATCGGCTCG 1419
QY 2554 CTCAGAGGAACAAGCGAGCGGTGTACACAAGGCTCCCTGTAGTTGGGACCTGGATCAAG 2613
DB 1420 CAGAGGTGAACAGCGCTGGGTGTACACCCGTGTACCTCTCTTGGACTGGATCCACG 1479
QY 2614 AGCACACTGGGGGTATAGC 2631
DB 1480 AGCAGATGGAGAGAGACC 1497

RESULT 11
US-09-518-046-3
; Sequence 3, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
US-09-518-046-3

Query Match 3.5%; Score 109.8; DB 3; Length 2544;
Best Local Similarity 54.6%; Pred. No. 2.3e-19;
Matches 266; Conservative 0; Mismatches 212; Indels 9; Gaps 2;

QY 2151 ATCATCACCCACCTTCTTCAATGATTTTCACTTCGACTATGACATCGCTTCTGGAG 2210
DB 1145 ATTGTCTACACAGCAAGTACAGCCAAAGAGGTGGGCAATGACATCGCCCTTATGAAG 1204
QY 2211 CTGGAAGTCTGGTGGAGTACAGCACCGTCTGTGGCCCTCATCTGCTGCTGTGTGTGTGT 2270
DB 1205 CTGGCCGGGCACTCAGCTTCAATGAAATGATCCAGCTGTGTGCTGCTGCCAATCTGAA 1264
QY 2271 CATGTCTTCTCTGCTGCAAGGCCATCTGGGTACAGGCTGGGGGCAACAAGAGAGGA 2330
DB 1265 GAGAACTTCCCGATGGAAGAAAGTGTGTGACGCTGAGGATGGGGGGCCACAGAGATGA 1324
QY 2331 GGTACCGAGCGGTGATCTCTGCAAGAGGGTGTGAGATCGGTGTATCAACAGACCATCTGT 2390
DB 1325 GGTACGCTTCCCTGTCTGTAACCGCGCGCGCTCCCTTTGATTTTCCACAGATCTGC 1384
QY 2391 GA-----GGACCTCATGTCGCGACAGATCAACCCAGAAATGATGTGTGGTTCCTC 2444
DB 1385 AACACAGGACGCTGTACGGTGGCATCATCTCCCCCTCCATGTCTGTGCGGGGTACCTG 1444
QY 2445 AGTGGGGGTGTGACCTCTCCAGGGGTGACTCTGTGGGGCCCCCTTGTCAAGCGGGAGAAA 2504
DB 1445 ACGGGTGGCGTGGACAGCTGCCAGGGGGGACAGCGGGGGCCCCCTGTGTGTCAAGAGAGG 1504

QY 2505 GATGGGCAATGTTCCAGGCTGGTGTGTGAGCTGGGTGAAGCTGGCTCAGAGAAC 2564
Db 1505 AGGCTGTGGAAGTT--AGTGGAGACGACCGCTTTGGCATCGCTGGCGAGAGTGAAC 1561
QY 2565 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGG 2624
Db 1562 AAGCCTGGGTGTACACCGGTGTACCTCCCTCTGGAAGTGGATCCACGAGCAGATGGAG 1621
QY 2625 GTATAGC 2631
Db 1622 AGAGACC 1628

RESULT 12

US-09-342-749-1

Sequence 1, Application US/09342749

Patent No. 6166194

GENERAL INFORMATION:

APPLICANT: Worg, Alexander K.C.

APPLICANT: Tavtigian, Sean V.

APPLICANT: Teoq, David H.-F.

APPLICANT: Myriad Genetics, Inc.

TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor

FILE REFERENCE: 2318-202

CURRENT APPLICATION NUMBER: US/09/342,749

CURRENT FILING DATE: 1999-06-29

EARLIER APPLICATION NUMBER: US 60/091,044

EARLIER FILING DATE: 1998-06-29

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1479

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1476)

FEATURE:

NAME/KEY: conflict

LOCATION: (724)

OTHER INFORMATION: Listed as T in GenBank Accession No. U75329

FEATURE:

NAME/KEY: conflict

LOCATION: (985)

OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329

FEATURE:

NAME/KEY: conflict

LOCATION: (1347)

OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329

FEATURE:

NAME/KEY: conflict

LOCATION: (1466)

OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329

FEATURE:

NAME/KEY: conflict

LOCATION: (1471)

OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.

FEATURE:

NAME/KEY: allele

LOCATION: (478)

OTHER INFORMATION: This base can be G or A with G being the more

OTHER INFORMATION: common allele. The codon will change from Val to

FEATURE:

NAME/KEY: allele

LOCATION: (777)

OTHER INFORMATION: This base can be C or T with C being the more

OTHER INFORMATION: common allele. The codon is unaffected with both

FEATURE:

NAME/KEY: allele

LOCATION: (768)

FEATURE:

NAME/KEY: allele

LOCATION: (768)

FEATURE:

NAME/KEY: allele

LOCATION: (768)

RESULT 13

OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
US-09-342-749-1

Query Match

3.3%; Score 101.6; DB 3; Length 1479;

Best Local Similarity 52.8%; Pred. No. 2,7e-17;

Matches 396; Conservative 0; Mismatches 324; Indels 30; Gaps 7;

QY 1886 CTTTACCAACAGGCTCGCGTGGTGTGGTGGACGAATGCGGAGGAGGCGAGTGGCCCTG 1945
Db 747 CTCAAGCGCCAGAGAGGATCGTGGCGCGAGAGCGCGCTCCCGGGGCGCTGGCCCTG 806
QY 1946 GCAGGTGAGCTCCACGCCCTGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCC 2005
Db 807 GCAGGTGAGCTCCACGCCCTGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCC 863
QY 2006 TGACTGGCTGGTCTCTGCGAGCTCAITGCTTTAGGATGACAAAATTTCAAGTACTCAGA 2065
Db 864 CGAGTGGATCGTACAGCGCGCCCACTGC-----GTGGAAAAACCTCTTAAACAATCCATG 917
QY 2066 CTACAGGATGTGAGCGGCTTCTGGTCTGTGGACCAAGCGAGCGAGTGGCTCTGG 2125
Db 918 GCAT-----TGGACGGCATTTGCGGGGATTTGAGA-----CAATCTTCATGTTCTAT 966
QY 2126 GTGCGAGGAGCTGAAGCTCAACGATCATCACCCACCTTCTCTCAATGATTTCACTT 2185
Db 967 GGAGCGGATACCAAG-TAGAAAAAGTGATTTCTCATCCAAATTAATGACTCCCAAGACCA 1025
QY 2186 CGACTATGATCGCTTCTGAGAGTGGAGAGTGGTGGAGTACAGACCGTCTGGCG 2245
Db 1026 GAACAATGATTCGCTGATGAAGCTGCAGAGCGCTCTGACTTTTCAACGACCTAGTAA 1085
QY 2246 CCCCATCTGCTGCTGATCTACCATGCTTCCCTGCTGGCAAGCCCATCTGGGTCAC 2305
Db 1086 ACCAGTGTGCTGCCCAACCCAGGCGATGCTGCGAGCCAGACAGCTCTGCTGGATTC 1145
QY 2306 AGCTGGGGGACACAAAAGAGGAGGTA-----CCGAGCGCTGATCTTCGAGAAGGG 2359
Db 1146 CGGGTGGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAGTCTCTGAACGCTGCCAAGT 1205
QY 2360 TGAGATCCGCTGCTCATCAACAGACCACTGTGAGGACCTCATGCCCGCAGCATACCC 2419
Db 1206 GCTTCTCATGTAGACACAGAGATGCAACAGCAGATATGTATGACAACCTGATACACC 1265
QY 2420 ACGAATGATGTGTGGGTTTCTCTCAGTGGGGGTGTGGACTCTCTGCCAGGCTGACTCTG 2479
Db 1266 AGCATGATCTGTGCGGCTTCTCTGAGGGGAACTGCGATTTCTTGCAGGCTGACAGTGG 1325
QY 2480 TGGCCCTTGTCAAGCGCGAGAGATGGGCGAATGTTTCCAGGCTGGTGTGGTGAAGTGG 2539
Db 1326 AGGGCTCTGTCTCACTTCCAGAGAACAAATATCTG--GTGGCTGATAGGCGATACAGCTG 1382
QY 2540 GGCTGAAGCTGCGCTCAGAGGAAAGCGCGGTGTACACAGGCTCCCTGTAGTTCG 2599
Db 1383 GGGTCTGCTGTGCCAAAGCTTACAGCAGAGTGTACGGGAATGTGATGTATTCAC 1442
QY 2600 GGAAGTGAATCAAGAGACACACTGGGGTATA 2629
Db 1443 GGAAGTGAATTCAGACAAATGAGGCGAGA 1472

US-09-691-840-1
; Sequence 1, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtighian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
; NAME/KEY: conflict
; LOCATION: (724)
; OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
; NAME/KEY: conflict
; LOCATION: (1466)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
; NAME/KEY: conflict
; LOCATION: (1471)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
; NAME/KEY: allele
; LOCATION: (478)
; OTHER INFORMATION: This base can be G or A with G being the more
; OTHER INFORMATION: common allele. The codon will change from Val to
; NAME/KEY: allele
; LOCATION: (777)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. The codon is unaffected with both
; NAME/KEY: allele
; LOCATION: (768)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. This is a silent polymorphism.
; NAME/KEY: allele
; LOCATION: (834)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. This is a silent polymorphism.
; NAME/KEY: allele
; LOCATION: (625)
; OTHER INFORMATION: This base can be T or A with T being the more
; OTHER INFORMATION: common allele. The codon will change from Phe to Ile
US-09-691-840-1
Query Match 3.3%; Score 101.6; DB 4; Length 1479;
Best Local Similarity 52.8%; Pred. No. 2.7e-17;
Matches 396; Conservative 0; Mismatches 324; Indels 30; Gaps 7;
Qy 1886 CTTTACCAACAGGCTCGGTGGTTGGTGGCAGCATGCGACGAGGCGGAGTGGCCCTG 1945
Db 747 CTCAGCGCGCAGAGCAGTGGTGGGCGGAGAGCGCTCCCGGGGCGCTGGCCCTG 806
Qy 1946 GCAGGTGAGCCTCCACGCCCTGGGCGAGGCGCACTTGTGTGGGCGCTCATCTCTCC 2005

Db 807 GCAGGTGAGCCTCGACGTCCAGAAC---GTCCAGCTGTGGGAGGCTCCATCATCACCCC 863
Qy 2006 TGAAGTGGTGGTCTCTGCTGAGCTCAATGCTTTTTCAGATGACAAAAATTTCAAGTACTCAGA 2065
Db 864 CGAGTGGATCGTACAGCGCGCCACTGC-----GTGAAAAAATCTTTAACAATCCATG 917
Qy 2066 CTACAGATGTGACCGGCTTCTGCTGGTCTGCTGGACAGAGCAAGCGCAGTGGCTTGG 2125
Db 918 GCAT-----TGGACGGCAATTTGGGGGATTTTGA-----CAATCTTTCATGTTCTAT 966
Qy 2126 GGTGAGGAGTGAAGCTCAAACTATCATCACCCACCTCTCTTCAATGATTTCACTT 2185
Db 967 GGAGCCGGATACCAAG--TAGAAAAAGTGAATTTCTATCCAAATATGATCTCAAGACCAA 1025
Qy 2186 CGACTATGACATCGCTTGTGAGAGCTGGAGAACTCGGTGAGTACAGCAGCCTGCTGG 2245
Db 1026 GAACAATGACATTCGCTGATGAAGCTGCAGAGCCTCTGACTTTCAACGACCTAGTGAA 1085
Qy 2246 CCCATCTGCTGCTGATGCTACCCATGTTTCCCTGTGGCAAGGCCATCTGGGTAC 2305
Db 1086 ACCAGTGTGTCTGCCCAACCCAGGCATGCTGCAGCCAGAACAGCTCTGCTGGATTC 1145
Qy 2306 AGGCTGGGGGCACAAAAAGAGGAGGTA-----CCGGAGCGCTGATCTCTGCAGAGGG 2359
Db 1146 CGGTGGGGGGCCACCGAGGAGAAAGGAAAGCTCAGAGTGTCTGACCGTGCAGAGGT 1205
Qy 2360 TGAGATCGGTGTATCAACAGACCACTGTGAGGACCTCATGCCGAGCAGATCACCCC 2419
Db 1206 GCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACACC 1265
Qy 2420 ACNAATGATGTGTGGGTTTCTCAGTGGGGGTGTGACTCTCTGCCAGGCTGATCTGG 2479
Db 1266 AGCATGATCTGTGCCGCTTCTCTGCGAGGGGAACGTCGATTTTCCAGGCTGACGTTG 1325
Qy 2480 TGGCCCTCTGTAAGCGCGAGAAAGATGGCGAATGTTCCAGGCTGTGTGGTGGCTG 2539
Db 1326 AGGCTCTGTGTCTACTTTCGAGAAACAATATCTG---GTGGCTGATAGGGGATACAAGCTG 1382
Qy 2540 GGTGTAAGTGGCTCGCTCAGAGGAACAAGCCAGGCGGTGTACAAAGGCTCCCTGTAGTTCG 2599
Db 1383 GGGTCTGCTGTGCCAAAGCTTACAGCAGGAGTGTACGGGAATGTGATGGTATTTCAC 1442
Qy 2600 GGAAGTCAAGAGCAGACACTGGGTATA 2629
Db 1443 GGAAGTCAAGAGCAGACACTGGGTATA 1472

RESULT 14
US-09-016-366A-20
; Sequence 20, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: WAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998


```

;
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
;
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-016-366A-20
;
; Query Match 3.3%; Score 101.2; DB 2; Length 1128;
; Best Local Similarity 53.1%; Pred. No. 3.1e-17;
; Matches 394; Conservative 0; Mismatches 288; Indels 60; Gaps 6;
;
; QY 1903 GCGTGGTTGGTGGACGAATCGGACGAGGCGGAGTGGCCCTGGCAGTGGAGCC-----1956
; DB 87 GCATCGTTGGGGGTGAGAGGCCCCCAGGAGCAAGTGGCCCTGGCAGTGGAGCTGAGAG 146
;
; QY 1957 TCCAGGCCCTGGGCGAGGCCACTGTGTGGGGGCTCGCTCATCTCTCTGACTGGCTGG 2016
; DB 147 TCCAGGCCCTGATGATGCACTTCTGCGGGGCTCCCTCATCCACCCGAGTGGGTGC 206
;
; QY 2017 TCTCTGAGCTCATTTGCTTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGT 2076
; DB 207 TGACCGCAGCGCACTGGTGGGACCGGACGTCAA-----GG 242
;
; QY 2077 GGAGCGCTTCTCTGGGTCTCTGCTGGAACGAGCAAGCGAGTGGCTCTGGGGTGGAGAGC 2136
; DB 243 ATCTGGCGCCCTCAGGCTGCACTGGGGAGCAGCCTTACTACGAGCAGCTGC 302
;
; QY 2137 TGAAGCTCAACGATCATCAACCCGCTTCTTCAATGATTTCACTTGCATATGACA 2196
; DB 303 TGCCGGTCAGCAGGATCATCTGCAACCCACAGTTCTACACCGCCAGATCGGAGCGACA 362
;
; QY 2197 TCGCCTTCTGAGCTGAGAGTGGTGGAGTACAGCAGCGTGTGGCCCCCATCTGCC 2256
; DB 363 TCGCCCTGCTGGAGCTGGAGGAGCGGTGAAGGTTCTCCAGCCACGTCCACACGCTCACCC 422
;
; QY 2257 TGCCTGATGCTACCCATGCTTTCCTGCTGCAAGGCGCATCTGGGTACAGGCTGGGG-- 2314
; DB 423 TGCCCCCTGCTCAGAGACCTTCCCGCGGGATGCGGTGCTGGGTCACTGGCTGGGGG 482
;
; QY 2315 ---GCACACAAAGAGGAGGATACCGAGCGCTGATCTGCAAGAGGTGAGATCCGTG 2370
; DB 483 ATGTGGCAATGATGAGCGCTCCACCGCCATTTCTCTGAAGCAGGTGAAGTCCCCA 542
;
; QY 2371 TCATCAACACGACCACTGTGA-----GGACCTCATGCCGACGATC 2414
; DB 543 TAATGGAACCAATTTGTGACGAAATACACCTTGGCGCTTACCGGGAGACGAGC 602
;
; QY 2415 ACCCCAGCAATGATGTGT-----GTGGGTTTCTCAGTGGGGGTGAGTCTCTGCCAGG 2469
; DB 603 TCCGATCGTCCGTGACGACATGCTGTGTGCGGGAAACACCCCGAGGAGCTCATGTCCAGG 662
;
; QY 2470 GTGACTCTGGTGGCCCTTGTCAAGCGCGGAGAAAGATGGCGAATGTTTCAGGCTGGTG 2529
; DB 663 CGCACTCCGAGGCGCCCTGTGTGTGCAAGGTGA---ATGGCACCTGGGTGCGAGCGGGCG 719
;
; QY 2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGCAACAGCCAGGCGGTGTACACAGGCTCC 2589
; DB 720 TGGTCAGCTGGGGGAGGCGTGTGCCAGCCCAACCGGCTTGGCATCTACACCGGTGCA 779
;
;
; QY 2590 CTGTAGTTGGGACTGGATCAA 2611
; DB 780 CCTACTACTTGGACTGGATCCA 801
;
;
; RESULT 15
; US-08-978-404B-15
; Sequence 15, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-15
;
; Query Match 3.3%; Score 101.2; DB 2; Length 1128;
; Best Local Similarity 53.1%; Pred. No. 3.1e-17;
; Matches 394; Conservative 0; Mismatches 288; Indels 60; Gaps 6;
;
; QY 1903 GCGTGGTTGGTGGACGAATCGGACGAGGCGGAGTGGCCCTGGCAGTGGAGCC-----1956
; DB 87 GCATCGTTGGGGGTGAGAGGCCCCCAGGAGCAAGTGGCCCTGGCAGTGGAGCTGAGAG 146
;
; QY 1957 TCCAGGCCCTGGGCGAGGCCACTGTGTGGGGGCTCGCTCATCTCTCTGACTGGCTGG 2016
; DB 147 TCCAGGCCCTGATGATGCACTTCTGCGGGGCTCCCTCATCCACCCGAGTGGGTGC 206
;
; QY 2017 TCTCTGAGCTCATTTGCTTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGT 2076
; DB 207 TGACCGCAGCGCACTGGTGGGACCGGACGTCAA-----GG 242
;
; QY 2077 GGAGCGCTTCTCTGGGTCTCTGCTGGAACGAGCAAGCGAGTGGCTCTGGGGTGGAGAGC 2136
; DB 243 ATCTGGCGCCCTCAGGCTGCACTGGGGAGCAGCCTTACTACGAGCAGCTGC 302
;
; QY 2137 TGAAGCTCAACGATCATCAACCCGCTTCTTCAATGATTTCACTTGCATATGACA 2196
; DB 303 TGCCGGTCAGCAGGATCATCTGCAACCCACAGTTCTACACCGCCAGATCGGAGCGACA 362
;
; QY 2197 TCGCCTTCTGAGCTGAGAGTGGTGGAGTACAGCAGCGTGTGGCCCCCATCTGCC 2256
; DB 363 TCGCCCTGCTGGAGCTGGAGGAGCGGTGAAGGTTCTCCAGCCACGTCCACACGCTCACCC 422
;
; QY 2257 TGCCTGATGCTACCCATGCTTTCCTGCTGCAAGGCGCATCTGGGTACAGGCTGGGG-- 2314
; DB 423 TGCCCCCTGCTCAGAGACCTTCCCGCGGGATGCGGTGCTGGGTCACTGGCTGGGGG 482
;
; QY 2315 ---GCACACAAAGAGGAGGATACCGAGCGCTGATCTGCAAGAGGTGAGATCCGTG 2370
; DB 483 ATGTGGCAATGATGAGCGCTCCACCGCCATTTCTCTGAAGCAGGTGAAGTCCCCA 542
;
; QY 2371 TCATCAACACGACCACTGTGA-----GGACCTCATGCCGACGATC 2414
; DB 543 TAATGGAACCAATTTGTGACGAAATACACCTTGGCGCTTACCGGGAGACGAGC 602
;
; QY 2415 ACCCCAGCAATGATGTGT-----GTGGGTTTCTCAGTGGGGGTGAGTCTCTGCCAGG 2469
; DB 603 TCCGATCGTCCGTGACGACATGCTGTGTGCGGGAAACACCCCGAGGAGCTCATGTCCAGG 662
;
; QY 2470 GTGACTCTGGTGGCCCTTGTCAAGCGCGGAGAAAGATGGCGAATGTTTCAGGCTGGTG 2529
; DB 663 CGCACTCCGAGGCGCCCTGTGTGTGCAAGGTGA---ATGGCACCTGGGTGCGAGCGGGCG 719
;
; QY 2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGCAACAGCCAGGCGGTGTACACAGGCTCC 2589
; DB 720 TGGTCAGCTGGGGGAGGCGTGTGCCAGCCCAACCGGCTTGGCATCTACACCGGTGCA 779
;
;
; QY 2590 CTGTAGTTGGGACTGGATCAA 2611
; DB 780 CCTACTACTTGGACTGGATCCA 801
;
;
; RESULT 15
; US-08-978-404B-15
; Sequence 15, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-15

```

```
QY 2197 TCGCCTTGTGAGGTGAGAAAGTGGTGGAGTACAGCACCGTGTGTGGCCCCCATCTGCC 2256
Db |||||
QY 363 TCGCCTTGTGAGGTGAGAGAGCGGTGAAGGTCTCCAGCCACGTCCACACGGTCAACC 422
Db |||||
QY 2257 TGCCTGATGCTACCCATGCTTTCCCTGTGCAAGGCCATCTGGGTACAGGCTGGG-- 2314
Db |||||
QY 423 TGCCTTGTGCTCAGAGACCTTCCCTCCCGGGATGCCGTGGGTCACTGGCTGGGGCG 482
Db |||||
QY 2315 ----GCACACAAAGAGGAGGTACCGGAGCGCTGATCCTGCAGAAAGGTGAGATCCCGTG 2370
Db |||||
QY 483 ATGTGGACAAATGATGAGCGCCTCCACCGCCATTTCTCTGAAGCAGGTGAAGTCCCCCA 542
Db |||||
QY 2371 TCATCAACCCAGACCACTGTGA-----GGACCTCATGCCGAGCAGATC 2414
Db |||||
QY 543 TAATGGAACCCACATTTGTGACGCAAAATACCACTTGGCGCCTACACGGGAGACGACG 602
Db |||||
QY 2415 ACCCCACGAATGATGTGT-----GTGGGTTTCTCAGTGGGGGTGTGGACTCTGCCAGG 2469
Db |||||
QY 603 TCCGCATGTCGTGACGACATGCTGTGTGCCGGGAACACCCGAGGACTCATGCCAGG 662
Db |||||
QY 2470 GTGACTGTGGTGGCCCCCTTGTCAAGCGGAGAAAGATGGCGAATGTTCCAGGCTGGTG 2529
Db |||||
QY 663 GCGACTCCGGAGGGCCCCCTGGTGTCAAAGTGA---ATGGCACCTGGCTGCAGGGGGCG 719
Db |||||
QY 2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGGTGTACACAGGCTCC 2589
Db |||||
QY 720 TGGTCAGCTGGGGCGAGGGCTGTGCCAGCCCAACCGGCTGGCATCTACACCCGTTGCA 779
Db |||||
QY 2590 CTGTAGTTCGGGACTGGATCAA 2611
Db |||||
QY 780 CTTACTTGGACTGGATCCA 801
Db |||||
```

Search completed: February 21, 2004, 07:15:18
Job time : 193 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 05:28:33 ; Search time 1052 Seconds
(without alignments)
10875.821 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catggtagacggctgcgcgg.....ttaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3106	100.0	3106	9	US-09-900-751-1	Sequence 1, Appli
2	1893.8	60.8	3149	12	US-10-295-027-968	Sequence 968, Appli
3	1893.2	60.6	3147	11	US-09-776-191-1	Sequence 1, Appli
4	1893.2	60.6	3147	11	US-09-776-191-49	Sequence 49, Appli
5	1893.2	60.6	3147	12	US-10-147-211A-1	Sequence 1, Appli
6	1893.2	60.6	3147	12	US-10-147-211A-3	Sequence 3, Appli
7	1893.2	60.6	3147	12	US-10-156-214A-1	Sequence 1, Appli
8	1893.2	60.6	3147	13	US-10-190-030B-1	Sequence 1, Appli
9	1893.2	60.6	3147	13	US-10-190-030B-3	Sequence 3, Appli
10	1893.2	60.6	3147	13	US-10-267-219-1	Sequence 1, Appli
11	1893.2	60.6	3147	13	US-10-267-219-3	Sequence 3, Appli
12	1893.2	60.6	3147	13	US-10-302-840A-1	Sequence 1, Appli
13	1893.2	60.6	3147	13	US-10-302-840A-3	Sequence 3, Appli
14	1893.2	60.6	3147	13	US-10-112-221A-1	Sequence 1, Appli
15	1893.2	60.6	3147	13	US-10-112-221A-3	Sequence 3, Appli

16	1893.2	60.6	3147	13	US-10-104-271-1	Sequence 1, Appli
17	1893.2	60.6	3147	13	US-10-104-271-3	Sequence 3, Appli
18	1893.2	60.6	3147	15	US-10-099-700A-1	Sequence 1, Appli
19	1893.2	60.6	3147	15	US-10-099-700A-3	Sequence 3, Appli
20	1559.2	50.2	2756	9	US-09-925-301-351	Sequence 351, App
21	1532.4	49.3	3112	15	US-10-097-340-311	Sequence 311, App
22	556.4	17.9	726	15	US-10-092-004A-1	Sequence 1, Appli
23	328	10.6	472	11	US-09-318-995-35415	Sequence 35415, A
24	306.8	9.9	434	10	US-09-736-457-1480	Sequence 1480, Ap
25	306.8	9.9	434	10	US-09-902-941-1480	Sequence 1480, Ap
26	306.8	9.9	434	10	US-09-849-626-1480	Sequence 1480, Ap
27	306.8	9.9	434	13	US-10-113-872-1480	Sequence 1480, Ap
28	306.8	9.9	434	15	US-10-017-754-1480	Sequence 1480, Ap
29	302.6	9.7	429	10	US-09-867-701-5310	Sequence 5310, Ap
30	265	8.5	591	9	US-09-815-343-1503	Sequence 1503, Ap
31	246.8	7.9	393	10	US-09-960-352-5954	Sequence 5954, Ap
32	240.8	7.8	2672	11	US-09-776-191-9	Sequence 9, Appli
33	240.8	7.8	2672	12	US-10-156-214A-9	Sequence 9, Appli
34	240.8	7.8	3104	11	US-09-776-191-7	Sequence 7, Appli
35	240.8	7.8	3104	12	US-10-156-214A-7	Sequence 7, Appli
36	239.2	7.7	1327	10	US-09-378-295A-170	Sequence 170, App
37	239.2	7.7	1327	10	US-09-978-697-170	Sequence 170, App
38	239.2	7.7	1327	10	US-09-978-192A-170	Sequence 170, App
39	239.2	7.7	1327	10	US-09-999-832A-170	Sequence 170, App
40	239.2	7.7	1327	11	US-09-978-189-170	Sequence 170, App
41	239.2	7.7	1327	11	US-09-378-608A-170	Sequence 170, App
42	239.2	7.7	1327	11	US-09-378-585A-170	Sequence 170, App
43	239.2	7.7	1327	11	US-09-978-191A-170	Sequence 170, App
44	239.2	7.7	1327	11	US-09-978-403A-170	Sequence 170, App
45	239.2	7.7	1327	11	US-09-978-564A-170	Sequence 170, App

ALIGNMENTS

RESULT 1
US-09-900-751-1
; Sequence 1, Application US/09900751
; Patent No. US20020026653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900,751
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3106
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-900-751-1

Query Match 100.0%; Score 3106; DB 9; Length 3106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGGTAGACGGCTGCCCGGAGGACCAACGCGTGTGAGACCGCGATCGGACCGCCAAAA 60

Db 1 CATGGTAGACGGCTGCCCGGAGGACCAACGCGTGTGAGACCGCGATCGGACCGCCAAAA 60

Qy 61 CCATGGTAGCAATCGGGGCGGAGCCGAGCGGGGCTCTCAGGACTTCGGCCGGGAC 120

Db 61 CCATGGTAGCAATCGGGGCGGAGCCGAGCGGGGCTCTCAGGACTTCGGCCGGGAC 120

QY 121 TCAAGTACAACTCCCGCTAGAGAACTGATGGCTTTGAGAGGGTGTGGAGTTCTTGC 180
Db 121 TCAAGTACAACTCCCGCTAGAGAACTGATGGCTTTGAGAGGGTGTGGAGTTCTTGC 180
QY 181 CTGCGAACTAATCCCAAGAAAGTGGAGAAAGAGAGCCCGAGGGCTGGGTGGTGGTGG 240
Db 181 CTGCGAACTAATCCCAAGAAAGTGGAGAAAGAGAGCCCGAGGGCTGGGTGGTGGTGG 240
QY 241 CAGTCTGTTCAGTCT 300
Db 241 CAGTCTGTTCAGTCT 300
QY 301 ATTATCGGAATGTGGGTCTCAAAAGTCTCAATGGCACTGAGGATCAAAATGAGA 360
Db 301 ATTATCGGAATGTGGGTCTCAAAAGTCTCAATGGCACTGAGGATCAAAATGAGA 360
QY 361 TCTTTCTGGATGCGTATGAGAACTCCACTCCACAGAGTTTATAGCGCTGCCAGCCAGG 420
Db 361 TCTTTCTGGATGCGTATGAGAACTCCACTCCACAGAGTTTATAGCGCTGCCAGCCAGG 420
QY 421 TGAAGGAGCGTGAAGTGTGTGTAAGTCCCTGTCTGGGTCTCTCTCTCTCTCTCTCTCT 480
Db 421 TGAAGGAGCGTGAAGTGTGTGTAAGTCCCTGTCTGGGTCTCTCTCTCTCTCTCTCTCT 480
QY 481 AGTGGGTGTAACTGCTTCACTGAGGAGCTGTATCGCCCTACTACTGCTGAGGTTCA 540
Db 481 AGTGGGTGTAACTGCTTCACTGAGGAGCTGTATCGCCCTACTACTGCTGAGGTTCA 540
QY 541 GCATCCCCCACCACCTGGCAGAGAGTTGATCGCGCCATGGCTGTGAGGAGTTGTA 600
Db 541 GCATCCCCCACCACCTGGCAGAGAGTTGATCGCGCCATGGCTGTGAGGAGTTGTA 600
QY 601 CATTTGCCACCCGAGCAGCGGCTGTAATCTCTGCTGCTAACTGCTGAGTGGCTTCC 660
Db 601 CATTTGCCACCCGAGCAGCGGCTGTAATCTCTGCTGCTAACTGCTGAGTGGCTTCC 660
QY 661 CCAATTGACCCCGAGTGTGAGAGGACTCAGGACACAGCTGAGTGTGCTGAGTGTG 720
Db 661 CCAATTGACCCCGAGTGTGAGAGGACTCAGGACACAGCTGAGTGTGCTGAGTGTG 720
QY 721 CCCATGTCAGCAGTGTGACAGCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGG 780
Db 721 CCCATGTCAGCAGTGTGACAGCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGG 780
QY 781 CGCATGCCGCTGCTGAGTGGTCTCTGCGGGGAGCGCGACTGTGCTGAGCTCACT 840
Db 781 CGCATGCCGCTGCTGAGTGGTCTCTGCGGGGAGCGCGACTGTGCTGAGCTCACT 840
QY 841 TCCGAAGCTTTGATGTGCTCTCTCTGATGAGCATGSCAGTGCCTGCTCAGCTGTATG 900
Db 841 TCCGAAGCTTTGATGTGCTCTCTCTGATGAGCATGSCAGTGCCTGCTCAGCTGTATG 900
QY 901 ATAGCCTGAGCCCCATGGAACCCACAGCTGTGTGCGGCTGTGTGTCACCTTCTCACCT 960
Db 901 ATAGCCTGAGCCCCATGGAACCCACAGCTGTGTGCGGCTGTGTGTCACCTTCTCACCT 960
QY 961 CCTACACCTGACTTCT 1020
Db 961 CCTACACCTGACTTCT 1020
QY 1021 CTGACCGGCGACATCTCTGGCTTTGAGGCGCACTTCTTCTTCCAGCTGCCCAAGATGAGCT 1080
Db 1021 CTGACCGGCGACATCTCTGGCTTTGAGGCGCACTTCTTCTTCCAGCTGCCCAAGATGAGCT 1080
QY 1081 GTGGCGGCTTTTGTAGTGAACCCCAAGGACATTTAGAGCCCTTACTATCCAGGCCACT 1140
Db 1081 GTGGCGGCTTTTGTAGTGAACCCCAAGGACATTTAGAGCCCTTACTATCCAGGCCACT 1140
QY 1141 ACCGCGCCCAACATCACTGACATGGAATATCAGGTGCCCAACCAACCGGAACTGAGG 1200
Db 1141 ACCGCGCCCAACATCACTGACATGGAATATCAGGTGCCCAACCAACCGGAACTGAGG 1200
QY 1201 TGGCTTTCAAACTCTTCTATCTGTGTGACCCCAACAGTACCAGTGGGTCTCTGCAACAGG 1260

Db 1201 TGGCTTTCAAACTCTTCTATCTGTGTGACCCCAACAGTACCAGTGGGTCTCTGCAACAGG 1260
QY 1261 ACTATGTGGAGATCAACCGGGAGAAAGTACTCGGTGAGAGTCCAGTTTGTGTGAGCA 1320
Db 1261 ACTATGTGGAGATCAACCGGGAGAAAGTACTCGGTGAGAGTCCAGTTTGTGTGAGCA 1320
QY 1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCTCACTGATACAGGACACCGGGT 1380
Db 1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCTCACTGATACAGGACACCGGGT 1380
QY 1381 TCTTAGCTGATCTCTCTCTTAAGACTCCAAAGCCCGTGGCCAGGAGTGTTCATGTGCA 1440
Db 1381 TCTTAGCTGATCTCTCTCTTAAGACTCCAAAGCCCGTGGCCAGGAGTGTTCATGTGCA 1440
QY 1441 AGACTGAGCGGTGATCGGAAGAACTGCGCTGCGAGCGGTGGGAGAGTGTCCCGGAT 1500
Db 1441 AGACTGAGCGGTGATCGGAAGAACTGCGCTGCGAGCGGTGGGAGAGTGTCCCGGAT 1500
QY 1501 ATAGTGTAGCGGTACTTGGCGATGCAATGCCACCCACAGTTCACGTGCAAAACAGT 1560
Db 1501 ATAGTGTAGCGGTACTTGGCGATGCAATGCCACCCACAGTTCACGTGCAAAACAGT 1560
QY 1561 TCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGAGCGGAAGTGAAG 1620
Db 1561 TCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGAGCGGAAGTGAAG 1620
QY 1621 AGGAGGCTGCGAGTGTCTGCTGGGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTC 1680
Db 1621 AGGAGGCTGCGAGTGTCTGCTGGGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTC 1680
QY 1681 AGAGCCAGAAAGTGTAAATGGGAAGCAACTGTGGAGATGGTCTGACGAGGCTTCATGTG 1740
Db 1681 AGAGCCAGAAAGTGTAAATGGGAAGCAACTGTGGAGATGGTCTGACGAGGCTTCATGTG 1740
QY 1741 ACAGCTGATGTCTCTCTGTCACAAATATACCTACCGCTGCAAAATGGGCTCTGTCTC 1800
Db 1741 ACAGCTGATGTCTCTCTGTCACAAATATACCTACCGCTGCAAAATGGGCTCTGTCTC 1800
QY 1801 TGAGCAAGGCAACCCCTGAGTGTGTGGGAAGAGCGACTGTAGGATGGCTCCGATAGA 1860
Db 1801 TGAGCAAGGCAACCCCTGAGTGTGTGGGAAGAGCGACTGTAGGATGGCTCCGATAGA 1860
QY 1861 AAAATGTGACTGTGGGCTGCGATCTTTTACAAACAGGCTCGCTGTTGGTGGCACGA 1920
Db 1861 AAAATGTGACTGTGGGCTGCGATCTTTTACAAACAGGCTCGCTGTTGGTGGCACGA 1920
QY 1921 ATCGGACAGGCGAGTGGCCCTGGAGGTGAGCTTCCACGCCCTGGGCGAGGGCCACT 1980
Db 1921 ATCGGACAGGCGAGTGGCCCTGGAGGTGAGCTTCCACGCCCTGGGCGAGGGCCACT 1980
QY 1981 TGTGTGGGCGCTCGCTCATCTCTCTGACTGGTGTCTCTGAGCTCATTTGCTTTCAGG 2040
Db 1981 TGTGTGGGCGCTCGCTCATCTCTCTGAGCTGGTGTCTCTGAGCTCATTTGCTTTCAGG 2040
QY 2041 ATGACAAAATTTCAAGTACTCAGACTTACAGATGTGGAGCGGCTTCTGGGCTGTGTGG 2100
Db 2041 ATGACAAAATTTCAAGTACTCAGACTTACAGATGTGGAGCGGCTTCTGGGCTGTGTGG 2100
QY 2101 ACCAGAGCAGCGAGTGTCTTGGGGTGGAGAGCTGAGCTCAACAGTATCATCACCC 2160
Db 2101 ACCAGAGCAGCGAGTGTCTTGGGGTGGAGAGCTGAGCTCAACAGTATCATCACCC 2160
QY 2161 ACCCTTCTCTCAATGATTTTCACTTTCGATGATGATGCTTGGAGCTGGAGAGT 2220
Db 2161 ACCCTTCTCTCAATGATTTTCACTTTCGATGATGATGCTTGGAGCTGGAGAGT 2220
QY 2221 CGGTGAGTACAGCACTGTGGTGGCCCATCTGCTGCTGATGATGATGATGATGATGATG 2280
Db 2221 CGGTGAGTACAGCACTGTGGTGGCCCATCTGCTGCTGATGATGATGATGATGATGATG 2280
QY 2281 CTGCTGCAAGGCCATCTGGGTACAGGCTGGGGCAGCAAAAGAGGAGGTACCGGAG 2340

Db 2281 CTGCTGCAAGCCCATCTGGGTACAGCTGGGGGCACACAAAAGAGGAGGTACCGAG 2340
QY 2341 CGCTGATCTCGAAGGGTCAGATCGTGTATCAACACGACACACCTGTGAGACCTCA 2400
Db 2341 CGCTGATCTCGAAGGGTCAGATCGTGTATCAACACGACACACCTGTGAGACCTCA 2400
QY 2401 TGGCGCAGCAGATCACCCCAAGATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
Db 2401 TGGCGCAGCAGATCACCCCAAGATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
QY 2461 CTTGCCAGGGTACTCTGTGGCCCTTGTCAAGCGCGGAAAGATGGCGAATGTTC 2520
Db 2461 CTTGCCAGGGTACTCTGTGGCCCTTGTCAAGCGCGGAAAGATGGCGAATGTTC 2520
QY 2521 AGGCTGTGTGTGCTGAGCTGGGTGAGCGTGCCTCAGAGGAAACAGCCAGGCGTGTACA 2580
Db 2521 AGGCTGTGTGTGAGCTGGGTGAGCGTGCCTCAGAGGAAACAGCCAGGCGTGTACA 2580
QY 2581 CAAGGCTCCCTGTAGTTCGGGACTTGGATCAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
Db 2581 CAAGGCTCCCTGTAGTTCGGGACTTGGATCAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
QY 2641 AGCAGCGGACCAACACACGACAGGATGCCGACATGCATCACTGTGATACAGAGAG 2700
Db 2641 AGCAGCGGACCAACACACGACAGGATGCCGACATGCATCACTGTGATACAGAGAG 2700
QY 2701 GAACACTGACGACATTTATGCTGTGGCTCCCGCCCAACACACACGACGCTGTGAAC 2760
Db 2701 GAACACTGACGACATTTATGCTGTGGCTCCCGCCCAACACACGACGCTGTGAAC 2760
QY 2761 GCATCTTAGACTCAGAGTTCCTCAAGTGGGACCCCTCAAGAGTTCGAGAGAGAAC 2820
Db 2761 GCATCTTAGACTCAGAGTTCCTCAAGTGGGACCCCTCAAGAGTTCGAGAGAGAAC 2820
QY 2821 TTGCGTCTAGCGGCCAGCTGGGGCAGAGGTTTGAATGGCAGCCTCCCGCTTAGCC 2880
Db 2821 TTGCGTCTAGCGGCCAGCTGGGGCAGAGGTTTGAATGGCAGCCTCCCGCTTAGCC 2880
QY 2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTCTTCCAACTGTCAATGAGCTCCC 2940
Db 2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTCTTCCAACTGTCAATGAGCTCCC 2940
QY 2941 GGGAGCCTATGGGAGAGGGGTCAAGGTCATCTTTTTCAGGAAGCGCCAGCCCTAGGA 3000
Db 2941 GGGAGCCTATGGGAGAGGGGTCAAGGTCATCTTTTTCAGGAAGCGCCAGCCCTAGGA 3000
QY 3001 ACCCGAGAAAGAGTGTACTAAGCTGAATGTTTGTCTGTGCGAGGGGTGGGTAT 3060
Db 3001 ACCCGAGAAAGAGTGTACTAAGCTGAATGTTTGTCTGTGCGAGGGGTGGGTAT 3060
QY 3061 TTGAGAGTAAACATTTTATTTCTTTTAAAAAAGGGGGGGGGGGGGGGGGGGGGGG 3106
Db 3061 TTGAGAGTAAACATTTTATTTCTTTTAAAAAAGGGGGGGGGGGGGGGGGGGGGGG 3106

RESULT 2

US-10-295-027-968
; Sequence 968, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-01250005

; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 968
; LENGTH: 3149
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-968

Query Match 60.8%; Score 1889,8; DB 12; Length 3149;
Best Local Similarity 81.4%; Pred.No. 0;
Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;

QY 50 GACCGCCAAACCATGGGTAGCAATCGGGGCCCAAGCGCGAGGGGGCTCTCAGGACTT 109
Db 23 GGCCTCGGGACCATCGGAGCGATCGGCCCGCAAGGGCGGAGGGGCCCGAAGGACTT 82
QY 110 CGCGCGGGAGCTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGAGGGTGT 169
Db 83 CGCGCGGGAGCTCAAGTACAACTCCCGGCAACGAAAGTGAATGGCTTTGAGAGGAGCGT 142
QY 170 GGAATTCCTCGCTCGCAACAAATGCCAAGAAAGTGAAGAGCGAGCGCCAGCGCTGGGT 229
Db 143 GGAATTCCTCGCTCGCAACAAATGCCAAGAAAGTGAAGAGCGAGCGCCAGCGCTGGGT 202
QY 230 GGTGCTGGTGGCAGTCTGTTCAGTTCCTCTTGTCTCTCCCTCATGCTGGCTTGTCTGT 289
Db 203 GGTGCTGGCAGCGCTGTCTGATCGGCTCTCTTGTCTTGTCTGGGATCGGCTTCTGTGT 262
QY 290 GTGGCACTTCCATTATCGGAATGTGGGGTTCAAAGTCTTCAATGGCCATCTGAGGAT 349
Db 263 GTGGCACTTTCAGTACCGGGAGTGGTGTCCAGAAAGGCTTCAATGGCTTACATGAGGAT 322
QY 350 CACAAATGAGATCTTTCTGATCGGTATGAGAACTCCACTCCACAGAGTTTATCAGCCT 409
Db 323 CACAAATGAGATTTTGTGATGCTTACGAGAACTCCAACTCACTCAGTGTGTTTAAAGCCT 382
QY 410 GGCAGCCAGGTGAAGAGGCGCTGAAGCTGTGTACAATGAAGTCCCTGTCTCGGGTCC 469
Db 383 GGCAGCAAGGTGAAGAGCGCTGAAGCTGTGTACAGCGGAGTCCCATTCCTGGGCC 442
QY 470 CTACACACAGAGTGGCTGAAGTCTTCACTCCCTTCAAGTGGGCGAGTGTATCGCTACTCTG 529
Db 443 CTACACACAGAGTGGCTGTGAGCGGCTTTCAGCGGAGGAGCGCTCATCGCTACTACTG 502
QY 530 GTACAGATTCAGCATCCCGCCACACTCGCAGAGAGGTTGATCGCGCATGCTGTGGA 589
Db 503 GTCTGATTCAGCATCCCGCAGCAGCTGTGTGAGGAGGCGCGGCTCATGGCCAGGA 562
QY 590 GCGAGTGTAAATTCGCAACCCCGAGCAGCGGCACTGAAATCTTCTGCTGCTAACTCTGT 649

RESULT 3
US-09-776-191-1
; Sequence 1, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; OTHER INFORMATION: Nucleotide encoding MTSP1
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank AR081724
; PATENT DOCUMENT NUMBER: US Pat 5972616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-09-776-191-1
Query Match 60.6%; Score 1883.2; DB 11; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAACACCATGGGTAGCATCGGGCCGCAAGCCGCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCGGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 124
QY 165 GGTGTGAGTTCCTGCTCGGAACAAATCCAAAGAGTGGAGAAGCGAGGCCCGCCAGGGCGC 224
Db 125 GCGGTGGAGTTCCTGCGAGTCAACACGCTCAAGAGGTGGAAAGCATGGCCCGCGGGCGC 184
QY 225 TGGGTGGTCTGCTGGTGGAGTCTGTTCAGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGGTCTGCGAGCGGTGTGTATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGGTGTGGCACTTCCATTTATCGGAATGGCGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCACTTTCAGTACCGGACGTGGCTGTCCAGAGGTCTTCAATGGCTTACATG 304
QY 345 AGATACAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGATACAAAATGAGAAATTTTGTGATGCTTACGAGAACTCCAACTCCCACTGAGTTTGA 364
QY 405 AGCCTGGCCAGCCAGGTGAAGGCGGTGAAGGTCTGTATCAATGAAGTCCCTGTCTCTG 464

Db 365 AGCCTGGCCAGCAAGGTGAAGGCGGTGAAGCTGTCTGTATCAGCGGAGTCCCATTCCTG 424
QY 465 GGTTCCTTACCACAAAGAGTGGCTGTAACTGCTTCAAGTGAAGGCGAGTGTATCGCTAC 524
Db 425 GGGCCCTTACCACAAAGAGTGGCTGTAACTGCTTCAAGTGAAGGCGAGTGTATCGCTAC 484
QY 525 TACTGTGAGAGTTCAGCATCCCCACACCTGGGAGAGAGAGTGTATCGCGCCATGGCT 584
Db 485 TACTGTGAGAGTTCAGCATCCCCACACCTGGGAGAGAGAGTGTATCGCGCCATGGCT 544
QY 585 GTGAGCGAGTGTAACTATGCCACCCGAGCAGCGGCACTGAAATTCCTTCTGTCTTAA 644
Db 545 GAGGAGCGGTAGTATGCTGCCCCCGGGCGCTCCCTCAAGTCTTCTGTGGTCAAC 604
QY 645 TCTGTGTGGCTTCCCATTTGACCCAGAGTGTGAGAGGACTCAGAGCAACAGCTGC 704
Db 605 TCAGTGGTGGCTTCCCATTTGACCCAGGACTCCAAAACAGTACAGAGGAGCCAGCAACAGCTGC 664
QY 705 AGTTTTGGCCCTCATGCCCCCATGGTGCAGCAGTACACGCTTCACTACCCCTGGCTTCCC 764
Db 665 AGTTTTGGCTGCACGCCCCCGGTGTGAGCTGTGCGCTTCAACAGCCCGGCTTCCCT 724
QY 765 AACAGTCCCTTACCCGCGCATGCCCGCTGCCAGTGGGTCTCTCGGGGGAGCCCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCCCGCTGCCAGTGGGTCTCTCGGGGGAGCCCGACTCA 784
QY 825 GTGCTGAGCTTCACTTCCGAAGCTTGTATGCTGCTCTCTGTGATGAGATGGAGTGC 884
Db 785 GTGCTGAGCTTCACTTCCGAGCTTGTGACCTTGGCTCTCTGCGAGCGCGGCGAGCGAC 844
QY 885 CTGGTCAACGCTGTATGATAGCTGAGCCCATGGAACCCACGCTGTGTGGCTGGCTGTCT 944
Db 845 CTGGTCAACGCTGTATGATAGCTGAGCCCATGGAACCCACGCTGTGTGGCTGGCTGTCT 904
QY 945 GGCACCTTCTACCTCTCTACAACTGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTCTACCT 964
QY 1005 AGCTGTAAACCAATCTGACCGGCGCATCTGCTGGCTTGGAGCCACTTCTCTCAGCTG 1064
Db 965 ACACCTGTAAACCAATCTGACCGGCGCATCTGCTGGCTTGGAGCCACTTCTCTCAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CTTAGATGAGCAGCTGTGGAGCGCTTACGTAAGCCAGGAGCAATTTAGCAGCCCC 1084
QY 1125 TACTATCCAGGCGCACTACCCGCGCAACATCACTGCAATGGAATATCAAGTGGCCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCCGCGCAACATCACTGCAATGGAATATCAAGTGGCCCAAC 1144
QY 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGTGGAGCCCAACCGTACCAGTG 1244
Db 1145 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGTGGAGCCCAACCGTACCAGTG 1204
QY 1245 GGTCTCTGCACCAAGGACTATGTGAGATCAACGGGAGAGTACTGGGGTGGAGAGTCC 1304
Db 1205 GGTCTCTGCACCAAGGACTATGTGAGATCAACGGGAGAGTACTGGGGTGGAGAGTCC 1264
QY 1305 CAGTTTGTGGTGGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATCACTCG 1364
Db 1265 CAGTTTGTGGTGGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATCACTCG 1324
QY 1365 TACAGGAGACCGGTTCTTCTAGTGAAGTCTCTCTAGCTTCCCAAGCCCGTGGCCCA 1424
Db 1325 TACAGGAGACCGGTTCTTCTAGTGAAGTCTCTCTAGCTTCCCAAGCCCGTGGCCCA 1384
QY 1425 GGGATGTTTCAATGTGCAAGACTGGAGCTGCATCCGAAAGGAACTGCGCTGGAGCGGTGG 1484
Db 1385 GGGATGTTTCAATGTGCAAGACTGGAGCTGCATCCGAAAGGAACTGCGCTGGAGCGGTGG 1444
QY 1485 GCAGCTGCCCGGATTTAGTGAAGGCTTACTGCGCATGCAATGCCACCCAGCTTC 1544

Db 1445 GCCAGCTGCACCGACAGCAGCAGTGAAGTCACTCACTGAGTTCGACGCCGCCACCAAGTTC 1504
Qy 1545 ACCTGCAAAAACCAAGTTCGCAAGCCCCCTCTCTGCGGTCTGTGACAGTGTCAACAGACTGT 1604
Db 1505 ACCTGCAAGAACAAAGTTCGCAAGCCCCCTCTCTGCGGTCTGTGACAGTGTGAACAGACTGC 1564
Qy 1605 GGGACCGAAGTGTGACGAGGAGGCTGACGTCTCTGCTGGAGTTCCTCAAGTGTTCGAAT 1664
Db 1565 GGAACACACAGCAGCAGCAGGAGTGTGCTGCGGCCAGACCTTCAGGTGTTCGAAT 1624
Qy 1665 GGAAGTGTCTCTCCCTCAGCAGCAGCAGAGTGAATGGAAGGACCACTGTGAGAGTGGGTCT 1724
Db 1625 GGAAGTGTCTCTCGAAGGACAGCAGTGAATGGAAGGACCACTGTGAGAGTGGGTCT 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTCTCTGCGCCCAAGGTGAACCTGCTCACTTGTACCAACACACCTACCGCTGC 1744
Qy 1785 CAATATGCTCTCTGACGCAAGGCAACCTGAGTGTGATGGAAGACGAGTGTAGC 1844
Db 1745 CTCATGCGCTCTCTGAGCAAGGCAACCTGAGTGTGACGGAAGGAGGACTGTAGC 1804
Qy 1845 GATCGCTCCGATGAGAAAACTGTGACTGTGGCTGGATCTCTTACCAACAGGCTGCG 1904
Db 1805 GACGCTCAGATGAGAGGACTGCACTGTGGCTGGGTCACTTCAAGACAGGCTGCT 1864
Qy 1905 GTGTTGGTGCACGAATGCGACAGGCGCAAGTGGCTGGAGGTGAGCCTCCAGCC 1964
Db 1865 GTTGTGGGCGACCGATGCGGATGAGGCGAGTGGCTGGAGGTGAAGCTGCATGCT 1924
Qy 1965 CTGGGCGAGGCGCACTGTGTGGGCGCTCGCTCATCTCTGACTGCTGCTGTCTGCA 2024
Db 1925 CTGGGCGAGGCGCACTGTGGGCTCTCTCTCACTCTCTCCAACTGGCTGTCTGCG 1984
Qy 2025 GCTCATGCTCTTCAAGGATGACAAAAATTTCAAGTACTCAGACTACAGGATGAGGACGCGC 2084
Db 1985 GCACACTGCTACATGATGACAGAGGATTCAGGTACTCAGACCCACAGCAGTGAACGCGC 2044
Qy 2085 TTCTGCGGTCTGTGACGACAGGCAAGGCAAGTGCCTCTGGGCTGAGGAGCTGAAGCTC 2144
Db 2045 TTCTGCGGTCTGTGACGACGACGACGACGACGCGCGCTGGGCTGAGGAGCGCAGCTC 2104
Qy 2145 AAAGTATATATCAACCCCTCTCTCAATGATTTCACTTCACTTCACTGATGATGATGCTG 2204
Db 2105 AAGCGATATCTCCACCCCTCTCTCAATGATTTCACTTCACTTCACTGATGATGATGCTG 2164
Qy 2205 CTGAGCTGGAGAGTCTGGTGGATGACAGCAGCCTGCTGGCCCATCTGCTGCTGCTGAT 2264
Db 2165 CTGAGCTGGAGAAACCGCAGAGTACAGCTCCATGCTGGGCCCATCTGCTGCTGCGGAC 2224
Qy 2265 GCTACCATGCTCTCTCTGCTGGCAAGGCAATCTGGGTGACAGCTGGGCGACACAAAA 2324
Db 2225 GCTCCCATGCTCTCTCTGCTGGCAAGGCAATCTGGGTGACAGCTGGGCGACACACACAG 2284
Qy 2325 GAGGAGGTACCGGAGCGCTGATCCTGAGAGGCTGAGATCCGTGATCATCAACACAGACC 2384
Db 2285 TATGAGGCACTGCGGCTGATCTCTGAAAGGCTGAGATCCCGTATCATCAACACAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCGCAGCAGATCACCCACGAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAACCTCTCTGCGCAGCAGATCACCGCGCATGATGTGCTGGGTTCCTC 2404
Qy 2445 AGTGGGCTGTGAGTCTCTGCGAGGCTGACTCTGCTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGCGTGGACTCTCTGCGAGGCTGATTCGCGGGGACCCCTGTCCAGCGTGGAGCG 2464
Qy 2505 GATGGCGGAATGTTCCAGGCTGTGTGTGAGCTGGGCTGAGGCTGGCTCGAGGAAAC 2564
Db 2465 GATGGCGGAATCTTCCAGGCGCGTGTGTGAGCTGGGAGACGCGCTGCGCTCGAGGAAAC 2524
Qy 2565 AAGCAGCGGTGTACACAGGCTCCCTGTAGTGTGGGACTGGATCAAGAGACACACTGGG 2624
Db 2525 AAGCAGCGGTGTACACAGGCTCCCTGTGTTTGGGACTGGATCAAGAGAACACTGGG 2584

Qy 2625 GTATAGCAGCATGACAGACAGCGACCAAAACACCCACAGGATCCCGACATGCACA 2684
Db 2585 GTATAGGGGCGGG---GCCACCAATGTGTACACTGCGGGCCACCATCGTCCACC 2641
Qy 2685 CTTGATACAGGAGAGAACACTGACGACATTTATGTGTGGCTCCCCCCCCCAACACA 2744
Db 2642 CCAAGTGTGCAGC-CCTGAGGCTGGAGCTGGACCGCTGACTGCACACGCGCCC-CCAGA 2699
Qy 2745 ACCAGACTGTGAAGTGCATCTTATAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAAGTCAATCTCCAGGCTCCAAAT 2735

RESULT 4

US-09-776-191-49
; Sequence 49, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiunn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTP51
US-09-776-191-49

Query Match 60.6%; Score 1883.2; DB 11; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
Qy 45 GATCGGACCGCCAAACCATGGTAGCATCGGGCGCGAAGCCGAGGGGCTCTCAG 104
Db 5 GAGCGCGCTCGGGGTACCATGGGGAGCGATCGGGCGCCGCAAGGGCGAGGGGCGCGAAG 64
Qy 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 124
Qy 165 GGTGTGGAGTCTCTGCTCGAACAAATGCCAAGAAAGTGGAGAGGAGGCCCCAGGCGC 224
Db 125 GCGTGGAGTCTCTGCGGCTCAACACGTCACAGAGTGGAGAAAGCATGSCCGGGGCGC 184
Qy 225 TGGGTGTGTGTGGAGTGTCTGTTAGCTTCTCTGCTTCCCTCATGGCTGGGCTT 284
Db 185 TGGGTGTGTGTGGAGCGGCTGTGATCGGCTCTCTCTGCTTGTGGGATCGGCTTC 244
Qy 285 CTGCTGTGGCACTTCCATTATCGGAATGTGGGCTTCAAAAAGCTTCAATGGCCATCTG 344

Db 245 CTGGTGGCAATTTGGAGTACCGGGAGCGTGTGTCAGGAAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAATGAGATCTTTCTGGATGGTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCACAAATGAGAAATTTTGGATGGCTTACGAGAACTCCAACTCCACTGAGTTTGA 364
QY 405 AGCTTGGCCAGCAGGTGAAGGAGGCGTGAAGCTGTCTGTACATGAAGTCCCTGTCTG 464
Db 365 AGCTTGGCCAGCAGGTGAAGGAGGCGTGAAGCTGTCTGTACAGGAGGATCCCAATTCCTG 424
QY 465 GGTCTCCTACCAAGAAAGTGGCTGTAACTGCTTCAAGTGAAGGAGTGTATCGCCCTAC 524
Db 425 GGCCTTACCAAGAGATGCTGTGACGGCTTTCAGGAGGAGCGTGTATCGCTAC 484
QY 525 TACTGCTCAGATTCAGATCCCCACACCTGGCAGAGAGGTTGATCGCCCTATGGCT 584
Db 485 TACTGCTCAGATTCAGATCCCCACACCTGGTGGAGAGGCGCTCATGGCC 544
QY 585 GTGGAGCGAGTTGTAACTTGTCCACCCGAGCAGCGGCACTGAATCTTCTGTCTAACA 644
Db 545 GAGGAGCGGTAGTATGCTGCTCCCGCGGGCGCTCCCTGAAGTCTCTTTGTGTACC 604
QY 645 TCTGTGGTGGCTTCCCAATTCACCCCAAGATGCTGCAGAGGACTCAGACAAAGCTGC 704
Db 605 TCAGTGGTGGCTTTCACCGAGCTTCCAAAAGTACAGAGGAGCCAGGACAAAGCTGC 664
QY 705 AGTTTGGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCC 764
Db 665 AGCTTGGGCTGCAGCGCGCGGTGTGGAGCTGATGCGCTTCAACAGCGCCGGCTTCCCT 724
QY 765 AACAGTCCCTACCGCGCATCGCGCTGCAGTGGTCTCTCGGAGGAGCGCCAGCTCT 824
Db 725 GACGCGCCCTACCGCGCTCATCGCGCTGCCAGTGGCGCTTCAACAGCGCCGGCTTCCCT 784
QY 825 GTGCTGAGCTCACCTTCGAAAGCTTTGATGTCGCTCCTGTGATGAGCATGGCAGTGC 884
Db 785 GTGCTGAGCTCACCTTCGAGCTTTGACTTGGTCTGTGAGAGCGCGGAGCGAC 844
QY 885 CTGGTACCGTGTATGATAGCTGAGCCCAATGGAACCCCAAGCTGTGTGGTGGCTGTGT 944
Db 845 CTGGTGAAGCTGTACAAACACCTGAGCCCAATGAGAGCCCAAGCTGTGTGGTGTGT 904
QY 945 GGCACCTTCTACCTCTCAGCACTGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCT 964
QY 1005 AGCTGATAACCAATCTGACCGCGCATCTCTGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 ACCTGATAACCAACTGAGCGCGGCTATCCGCTTTGAGGCCACTTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGAGCTGTGGCGGCTTTTGTGATGACACCCCAAGGAGCATTTAGCAGCCCC 1124
Db 1025 CTTAGATGAGAGCTGTGGAGCGGCTTTAGTAAAGCCCAAGGAGCATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCCACTACCGCCCAACATCACTGCATGGAATCAAGTGGCCCAAC 1184
Db 1085 TACTATCCAGGCCACTACCGCCCAACATCACTGCATGGAATCAAGTGGCCCAAC 1144
QY 1185 AACCGGAACGTGAAGGTGGCTTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1244
Db 1145 AACCGAATGTGAAGTGAAGTTCAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1204
QY 1245 GCTCTCTGACCAAGGACTATGTGGAGATCAACGGGAGAGTACTCGCGTGAAGTTC 1304
Db 1205 GGCACCTGCGGAGAGCTAGTGGAGATCAATGGGAGAAATCTCGCGGAGAGGTTC 1264
QY 1305 CAGTTTGGTGGAGCAACAGCAGCAGAGATTAAGTTCAGTTCCTCTCTCTCTCTCTCTCT 1364
Db 1265 CAGTTTGGTGGAGCAACAGCAGCAGAGATTAAGTTCAGTTCCTCTCTCTCTCTCTCT 1324
QY 1365 TACAGGACACCGGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
Db 1325 TACACCGACACCGGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1384

QY 1425 GGGATGTTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAATCGCTGCGACGSGCTGG 1484
Db 1385 GGGCAATTTCACTGCGCGACGGGGCGGTGTATCCGGAAGAGCTGCGCTGTGTGCTGG 1444
QY 1485 GCAGACTCCCGGATTAATGATGATGAGCGTTTACTCCGATGCAATGCCACCCAGTTTC 1544
Db 1445 GCGGACTCACCGACACAGCGATGAGCTCAACTGAGTTGGAGCGCGGCCACCACTTC 1504
QY 1545 ACGTCAAAAACCAAGTCTGCAAGCCCTCTCTCTGGGTCTGTGACAGTGTCAAGCTGT 1604
Db 1505 ACGTCAAAAACCAAGTCTGCAAGCCCTCTCTCTGGGTCTGTGACAGTGTCAAGCTGT 1564
QY 1605 GGGAGCGGAAGTGAAGAGGAGGCTGCAAGCTGTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGAGAACACAGCAGCAGCGGCTGCAAGTGTCCGCGCCAGACCTTCAGTGTTCGAAT 1624
QY 1665 GGGAGTGTCTCTCTCAGAGCCAGAGTGTAAATGGAAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAGTGTCTCTCTGAAAAGCCAGCAGTGCATGGAAGGAGCAGCTGTGGGAGCGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGCAACCAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTTCCTGCCCCAAGTGAACGTGCTCTCTGTACCAAAACACACCTTACCGCTGC 1744
QY 1785 CAAATGGCCCTCTCTCTGAGCAAGGCGCAACCTGAGTGTGTGGAAGACGAGCTGTAGC 1844
Db 1745 CTCAATGGGCTCTGCTTGTAGCAAGGCGCAACCTGAGTGTGACGGGAGGAGCTGTAGC 1804
QY 1845 GATGGCTCGATGAGAAAACCTGTGTGGTGTGCAATCTTTTACCAACAGGCTCGC 1904
Db 1805 GACGCTCAGATGAGAGGACTGTGCACTGTGGGTGCGGTCAATTCAAGAGACGAGCTCGT 1864
QY 1905 GTGGTGTGGTGCAGCAATGCGGAGCGGCGAGTGGCCCTGCGAGGTGAGCTCCACGCC 1964
Db 1865 GTTGTGGGGGACAGATGCGGATGAGGGCGAGTGGCCCTGCGAGGTGAAGCTCGATGCT 1924
QY 1965 CTGGGCGAGGGCCACTTGTGTGGGGCCCTCGCTCATCTCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCGAGGGCCACATCTGCGGTCTTCCCTCATCTCTCCCACTGGCTGGTCTCTGCC 1984
QY 2025 GCTCAATGTCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGATGTGACGCGCC 2084
Db 1985 GCACACTGTACTCATCGATGACAGAGGATTCAGGTACTCAGACCCCAAGCAGTGGACGCGCC 2044
QY 2085 TTCTTGGTCTGTGGACCAAGCAGCAGCTGCTCTGGGTGCGAGGCTGAGCTGAGCTC 2144
Db 2045 TTCTTGGCTTGCAGACCAAGCCAGCGAGCGCCCTGGGGTGCAGAGCGCAGGCTC 2104
QY 2145 AAACGTATCATCACCCACCTTCTTCAATGATTTCACTTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTTCGACTATGACATCGCGCTG 2164
QY 2205 CTGGAGCTGGAGAGTGGTGGAGTACAGCACCTGTGTGGCCCTCTGCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGGTAAGCTTCAATGTGTGGCGCCCTCTGCTGCGGAC 2224
QY 2265 GCTACCCATGTCTTCTCTCTGCGCAAGCCATCTGGGTTCAGAGCTGGGGGACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCTCTGCGGCAAGCCATCTGGGTTCAGGGCTGGGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGCTGAGATCCGTTGATCAACACAGACC 2384
Db 2285 TATGGAGCACTGGCGCTGATCTCTGAAAAGGCTGAGATCCGCTCATCAACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGGACAGCATCAACCCCAAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGCGAGAACTCTCTGCGGACAGCATCAACCCGCGCATGTGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGAGCTCTTCTGCGAGGGTGAATCTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGCTGGACTCTCTGCGAGGGTGAATCTGGGGGACCCCTGTCTCAGGTGGAGCG 2464

```
QY 2505 GATGGCGAATGTTCCAGGCTGGTGTGGTGGGTTGAAGCTCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATGTTCCAGGCGGTTGGTGGTGGGAGACGGCTCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGACAAAGGCTCCCTGTAGTTGGGACTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGCGGTGACAAAGGCTCCCTGTAGTTGGGACTGGATCAAGAGCACTGGG 2584
QY 2625 GTATAGCAGCATGACAGACAGCCGACCAACAAACACCCACAGGGATGCCGACATGCACA 2684
Db 2585 GTATAGGGCCGGG--GCCACCAATGTGTACACTGCGGGCCACCAATGCTCCACC 2641
QY 2685 CCTGGATACAGAGAGGAACAACACTGACCAATTTATGTGTGGCTCCGCCCCCAACACA 2744
Db 2642 CCAGTGTGCAGC-CCTGCGAGCTGGAGACTGGACCGCTGACTGCACCAAGCGCCC-CCAGA 2699
QY 2745 ACCGAGCTGTGAAGTCACTCCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 5
US-10-147-211A-1
; Sequence 1, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Yeh, Jiunn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSp1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AF081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-147-211A-1

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGTAGCAATCGGGCGCCGAAAGCGGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGCGGCGAG 64

QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCGGACGAGAAAGTGAATGGCTTGAGAG 124

QY 165 GGTGTGGAGTTCCTGCTCCGCAACATCCAGAAAGTGGAGAGCGAGGCGCCAGCGGC 224
Db 125 GCGGTGGAGTTCCTGCCAGTCAACAACTCAAGAGGTGGAAAGCATGGCCCGGGGCGC 184

QY 225 TGGGTGGTGGTGGGAGTGTGTGAGTTCCTTGTCTCTCCCTCATGGCTGGCTTG 284
Db 185 TGGGTGGTGGTGGGAGCGGTGCTGATCGGCTCTCTTGGTCTTGGTGGGATCGGCTTC 244
```

```
QY 285 CTGCTGTGGCACTTCCATTTATCGGAATGTGGGTTCAAAAAGTCTTTCAATGGCCATCTG 344
Db 245 CTGCTGTGGCACTTTCAGTACCGGACGTCGCTGCAGAGGCTTTCATGGCTACATG 304
QY 345 AGGATCAAAATGAGATCTTTCTGGATGCGGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGAAATTTTGGATGCGCTACGAGAACTCCAACTCCACTGAGTTGTA 364
QY 405 AGCTGSCCAGCAGGCTGAAGGAGGCGCTCAAGCTGTGTACAAATGAAGTCCCTGTCTCTG 464
Db 365 AGCTGSCCAGCAGGCTGAAGGAGGCGCTCAAGCTGTGTACAGCGGAGTCCCATTTCTCTG 424
QY 465 GGTCCCTTACCAAGAAGTCTGGCTGTAACTGCTTTCAGTGTAGGCGCAGTGTATCGCCCTAC 524
Db 425 GGTCCCTTACCAAGAAGTCTGGCTGTGACGCGCTTCAGCGAGGCGCAGCTCATCGCTAC 484
QY 525 TACTGTGAGATTCAGATTCGCCCACTGACCTGGCAGAGAGGTTGATCGGCGCATGGCT 584
Db 485 TACTGTGCTGAGTTTCAGCATCCCGCAGCACCTGTGTGGAGGCGCGAGCGCTCATGGCC 544
QY 585 GTGAGCGAGTGTGTAACTTGCACCCCGCAGCAGCGGCACTGAAATCCTTTCTGTCTAAACA 644
Db 545 GAGGAGCGGCTAGTGTATGCTGCCCCCGCGGCGGCTCTCTGAGTCTCTTTTGTGTCTACC 604
QY 645 TCTGTGTGGTGGCTTCCCTTACCCCAATTGACCCCAAGATGTGTGAGAGGACTCAGGACAAACAGCTGC 704
Db 605 TCAGTGTGTGGCTTCCCTTACCCCACTCCAAAACAGTACAGAGGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTTCCTGTGATGCCCATGTGTGAGCAGTGCACAGCTTCTACTACCTTGGCTTCCCC 764
Db 665 AGCTTTGGCTGTGACGCGCGGCTGTGTGAGCTGTATGCGCTTCCACGCGCGGCTTCCCT 724
QY 765 AACAGTCCCTTACCCCGCGCATGCCGTGCGAGTGGGTCTCTCGGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCCGTGCGAGTGGGCCCTTGGGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCCTCACTTCCGAAAGCTTTGATGTGCTCTCTCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACTTCCGAGCTTTGACTTGTGCTGCTGCGACGAGCGCGGCGAGCGAC 844
QY 885 CTGCTGACCGGTGTATGATGAGCTGAGCCCATGGAAACCCCAAGCTGTGTGGCTGTGT 944
Db 845 CTGCTGACCGGTGTATGACACACCCCTGAGCCCATGGAGCCCAAGCGCTGTGTGTGTGT 904
QY 945 GGCACCTTCTCAGCCTCTCAACACTGACTTTCCTCTCTCCAGAAAGCTTCTCTTGTCTC 1004
Db 905 GGCACCTTACCTTCCCTCTTACACTGACTTCCACTCTCTCCAGAAAGCTTCTCTCTCATC 964
QY 1005 AGCTGATTAACCAATCTGACCGGCGGACATCTGTGGCTTTTGGAGGCACTTTCTTCCAGCTG 1064
Db 965 ACATGATAACCAACATGAGCGCGGCACTCCCGGCTTTTGGAGGCACTTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGGTGTGACACCCCAAGGAGCATTTAGCAGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGGCGGCTTACGTAAAGCCCGAGGAGCACTTCAACAGCCCC 1084
QY 1125 TACTATCAGGCGCACTACCCGCCCAACATCAACTGACATGCAATATCAAGAGTGGCCAAAC 1184
Db 1085 TACTATCAGGCGCACTACCCCAACCAATGACTGTGCACATGGAACATTTAGAGTGGCCCAAC 1144
QY 1185 AACGGGAAGTGAAGTGGCTTCAAACTCTTCTATCTGTGTGAGACCCCAAGCTACCAAGTG 1244
Db 1145 AACGAGATGTGAAGTGAAGTCTTCAAACTCTTCTATCTGTGTGAGCGCGGCTGTGCTGCG 1204
QY 1245 GGCTCTGTGACCAAGGACTATGTGGAGATCAACGGGGAGAGTACTCGCGTGTGAGAGTCC 1304
Db 1205 GGCACCTTGGCCCAAGGACTACGTGAGATCAATGGGAGAGAACTACTCGGAGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTCCCACTTCCATTTCTGTACTCTG 1364
Db 1265 CAGTTTGTGTGAGCAGCAACAGCAACAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
```

QY 1365 TACACGACACGGGTTCTTAGTACGACCTCTCTACGACTCAACGACCCGCTGCCA 1424
DB 1325 TACACGACACCGGTTCTTAGTCAATACCTCTCTACGACTCCAGTACCCATGCCG 1384
QY 1425 GGGATGTCATGTGCAAGCTGAGCGGTGCATCCGAAAGGAATCTCGCTGCGACGGCTGG 1484
DB 1385 GGGCGATTCATGTGCGCAGCGGGGGGGTGTATCCGGAAGGAGCTCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTCCCGGATTATAGTATGATGAGGCTTACTCGCGATGCAATGCCACCCAGTTC 1544
DB 1445 GCCGACTGACCGACACAGCGATGAGCTCACTGAGTTGGAGCGCGCCAGCTTC 1504
QY 1545 ACGTCAAAAACCACTCTGCAAGCCCTCTTCTCGGCTCTGTGACAGTGTCAACGACTGT 1604
DB 1505 ACGTCAAGAAACAAGTCTGCAAGCCCTCTTCTCGGCTCTGTGACAGTGTGAACGACTGC 1564
QY 1605 GGGAGCGAAGTACGAGGCGGCTGACGCTGCTCTGCTGGAGTTTCAAGTGTCCAAAT 1664
DB 1565 GAGACAAACAGCGACGAGCGGGGTGCAGTTGTCCGGGCCAGACCTTCAGGTTGTTCCAAAT 1624
QY 1665 GGGAAAGTGTCTCCCTCAGAGCCAGAAAGTGTATGGAAAGGACAACTGTGGAGATGGGTCT 1724
DB 1625 GGGAAAGTGTCTCCGAAAGCCAGCAGTGCATGGGAGGAGCACTGTGGGACCGGTTCC 1684
QY 1725 GACGAGGTTCAATGTGACAGCGTGAATGCTCTTGTGACCAAAATATACCTACCGCTGC 1784
DB 1685 GACGAGGCTCTGTCGCCCAAGGTGAACGCTGCTCACTTGTATCCAAACACACCTTACCGCTGC 1744
QY 1785 CAAAATGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGATGGAGACGACGCTGTAGC 1844
DB 1745 CTCATGGGCTCTGTGTAGCAAGGCAACCTGAGTGTGACGGAAAGGAGACTGTAGC 1804
QY 1845 GATGCTCCGATGAGAAACCTGTGACTGTGGGCTGCGATCTCTTACCAAAACAGGCTCGC 1904
DB 1805 GACGCTCAGATGAGAGACTGCGACTGTGGGCTGCGGTCACTTACGAGACAGGCTGT 1864
QY 1905 GTGGTTGTGGCAAGATGCGGACAGGCGAGTGGCCCTGTGACGAGTGTGAGCTTCCACGCC 1964
DB 1865 GTTGTGGGGGACGAGTGTGCGGATGAGGCGAGTGGCCCTGTGACGAGTAAAGCTGATGT 1924
QY 1965 CTGGGCGAGGGCACTGTGTGGGCGCTCGCTCATCTCTCTGACTGTGCTGTCTGCA 2024
DB 1925 CTGGGCGAGGGCACTGTGGGCTGCTTCCCTCACTCTCTCCAACTGGCTGTCTGCTGCC 1984
QY 2025 GCTCAATGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
DB 1985 GCACACTGCTACATGATGACAGAGATTCAGTACTCAGACCCACGACGATGGAGGCC 2044
QY 2085 TTCCTGGGTCTGCTGGACAGAGCAAGCGCACTGTCTTGGGGTGGAGAGCTGAAGCTC 2144
DB 2045 TTCCTGGGCTTGCACGACCCAGAGCCAGCGCGCCCTTGGGGTGCAGAGCGCAGGCTC 2104
QY 2145 AAAAGTATCATCACCCCTTCTTCAATGATTTCACTTCCGATGTGACATGACATCGCCTTG 2204
DB 2105 AAGCGCATCATCTCCACCTTCTTCAATGACTTCACTTCCGATGTGACATCGCCTTG 2164
QY 2205 CTGGAGCTGGAGAGTTCGGTGGAGTACAGCAGCGTGTGCGGCCCATCTGCTGCTGCTGAT 2264
DB 2165 CTGGAGCTGGAGAACCGGACAGTACAGTCCATGTGTGCGGCCCATCTGCTGCTGCGGAC 2224
QY 2265 GCTACCATGTCTTCTGCTGCGAAGCGCATCTGGGTCAAGGCTGGGGGCAACAAAA 2324
DB 2225 GCTTCCCATGTCTTCTGCTGCGGACAGGCGCATCTGGGTCAAGGCTGGGGGCAACACCCAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGGTGTGATCGGTGTCTATCAACAGACC 2384
DB 2285 ATGAGGACCTTGGGCGCTGATCTCTGAAAAGGGTGTGATCCGCTGTATCAACAGACC 2344
QY 2385 ACGTGTGAGGACCTCATGCGCGAGCAGATCAACCCACGAATGATGTGTGGGTTTCTTC 2444
DB 2345 ACGTGGAGAACTCTCTGCGGACGAGATCAAGCCCGCATGATGTGTGGGCTTCTTC 2404
QY 2445 AGTGGGGGTGTGGAATCTCTGCGCAGGGGTGACTGTGGTGGCCCCCTTGTCAAGCGCGAGAA 2504

DB 2405 AGCGCGCGTGGAGCTCTGCTCCAGGGTATTTCGCGGGACCCCTCTCCAGCGTGGAGGG 2464
QY 2505 GATGGGCAATGTTTCCAGGCTGTGTGTGAGCTGGGTGAGGCTGCTCAGAGGAC 2564
DB 2465 GATGGGCGGATCTTCCAGGCGGTGTGTGTGAGCTGGGTGAGGACGGCTCGCTCAGAGGAC 2524
QY 2565 AAGCCAGCGGTGTACACAAAGCTCCCTGTAGTTGGGACTGATCAAAAGACACACTGG 2624
DB 2525 AAGCCAGCGGTGTACAAAGCTCCCTGTGTGGGACTGATCAAAAGACACACTGG 2584
QY 2625 GTATAGCAGTGTGACAGACAGCAGCCACCAAAACCCACAGGGATGCCGACATGCACA 2684
DB 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACTGCGGGGCCACCCATCGTCCACC 2641
QY 2685 CCTGATACAGAGAGGAGAACTGACGACATTTATGCTGTGCTCCCTCCCAACACA 2744
DB 2642 CAGTGTGACG-CCTGAGGCTGGAGACTGGACGCTGACTGCACACAGCGCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAATGATGATCTTTAGGACTCAGAGT 2780
DB 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 6
US-10-147-211A-3
; Sequence 3, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edward
; APPLICANT: Yeh, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1:
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-147-211A-3

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGACCGCCAAACCATGGGTAGCAATCGGGCCCGCAAGCGCGAGGGGCTCTCAG 104
DB 5 GAGCGGCTTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCCGAAG 64
QY 105 GACTTCGCGCGGGACTCAAGTACAACTCCCGGCTAGGACATCAATGAATGGCTTTGAGGAG 164
DB 65 GACTTCGCGCGGGGACTCAAGTACAACTCCCGGACACGAGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGAGATTTCTGCTGCTCGCAACAAATGCCAAGAAAGTGGAGAACGAGGCCCGCCAGGCGC 224
DB 125 GCGGTGGAGTTCTTCGCCAGTCAAACAGTCAAGAGGTGGAAAGCATGGCCCGGGGCGC 184
QY 225 TGGGTGTGCTGGTGGCAGTGTCTGTTCAGTTCCTCTTTGCTCTCTCCTCATGCTGCTGCTT 284
DB 185 TGGGTGTGCTGGCAGCGCTGTATCGGCTCTCTTTGCTCTTGTGGGATCGGCTTC 244
QY 285 CTGGTGTGGCACTTCCATATCGGAAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
DB 245 CTGGTGTGGCATTTGCACTACCGGACGTCGTGTGCCAGAGGTCTTCAATGGCTACATG 304

QY 345 AGGATCAAAATGAGATCTTTCTGGATCGGTATGAGAACTCCACTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGAAATTTTGTGGATCGCTTACGAGAACTCCAATCCACTGAGTTTGTGA 364
QY 405 AGCTGGCCAGCCAGGTGAAGGAGCGGTGAAGCTGTGTGAATGAAGTCCCTGTCTGTG 464
Db 365 AGCTGGCCAGCAGGTGAAGGAGCGGTGAAGCTGTGTGAATGAAGTCCCTGTCTGTG 424
QY 465 GGTCTCTACCAACAAGATCGGTGTAACTGCTTCACTGAGGGGAGGTGATCGCCTTAC 524
Db 425 GGCCCTTACCAACAAGATCGGTGTAACTGCTTCACTGAGGGGAGGTGATCGCCTTAC 484
QY 525 TACTGGTCAGAGTTTCAAGTATCCCCCAACCTTGGCAGAGAGGTGTGATCGGCCCATGGCT 584
Db 485 TACTGGTCAGAGTTTCAAGTATCCCCCAACCTTGGCAGAGAGGTGATCGGCCCATGGCT 544
QY 585 GTGGAGCAGATGTAACTTGGCAGCCCGAGCAGCGGCACTGAATCTTCTGTGTCTAAACA 644
Db 545 GAGGAGCGGTAGTATGCTGCTCCCGCGGGCGGCTCCCTGAAATCTCTTGTGTCTACC 604
QY 645 TCTGTGTGGCTTCCCAATGACCCCAAGATGCTGCAGAGGACTTCAGGACAAACAGCTGC 704
Db 605 TCAGTGTGTGGCTTCCCAAGGACTTCCAAACAGTACAGAGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTGGCTTGCATGCGCCATGGTGCAGCAGTGAACAGCTTCACTTACCCCTGGCTTCCCC 764
Db 665 AGCTTGGCTTGCATGCGCCCGGGTGTGAGCTGATGCGCTTCAACACGCCCGCTTCCCT 724
QY 765 AACAGTCTTACCCCGCGATGCGCGCTGCGAGTGGTCTGCGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCGCGCTGCCAGTGGGCGCTTCCAGAGCGTCTGCTGCTCA 784
QY 825 GTGTGAGCTTCACTTCCGAAGCTTTCATGTCGCTTCCCTGTGATGAGTACAGTGCAGTGC 884
Db 785 GTGTGAGCTTCACTTCCGAAGCTTTCATGTCGCTTCCCTGTGATGAGTACAGTGCAGTGC 844
QY 885 CTGTGTCACCGGTATGATGAGCTGAGCCCAATGGAACCCCAAGCTGTGTGGCGGTGTGT 944
Db 845 CTGTGTCACCGGTATGATGAGCTGAGCCCAATGGAACCCCAAGCTGTGTGGCGGTGTGT 904
QY 945 GGCACTTCTTCACTTCTTACAGTGTCTTCTCTCTCCAGAAAGTCTTCTCTGTC 1004
Db 905 GGCACTTCTTCACTTCTTACAGTGTCTTCTCTCTCCAGAAAGTCTTCTCTGTC 964
QY 1005 ACCTGATAACCAATACCTGACCGCGACATCTGCTTGTGAGCCACTTCTTCCAGCTG 1064
Db 965 ACCTGATAACCAATACCTGACCGCGCGCATCCGCGCTTGTAGCCCACTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGCAGTGTGGCGCTTTTGTGAGTGACA CCCAAGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGTGTGGAGCGCTTACGTAAGCCCGAGGGACATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCCACTACCCCGCCCAACATCACTGCATGAAATATCAAGTGCCTAAC 1184
Db 1085 TACTATCCAGGCCACTACCCCGCCCAACATGACTGCATGAAATATCAAGTGCCTAAC 1144
QY 1185 AACCGGAACGTGAAGGTGGCTTCAAACTCTTCTATCTGTGTGACCCCAAGTACAGTG 1244
Db 1145 AACAGCATGTGAAGGTGAGCTTCAAAATCTTCTACCTGCTTGAGCCCGGGGTGCTGG 1204
QY 1245 GCTCTCTGACCAAGGACTATGTGAGATCAACGGGGAGAGTATCGGGGTGAGAGTCC 1304
Db 1205 GGCACCTTGGCCCAAGGACTATCGTGGAGATCAATGGGGGAAATACTGCGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAGATTTACAGTCCACTTCCATCTGTATCACTCG 1364
Db 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAGATTTACAGTCCACTTCCATCTGTATCACTCG 1324
QY 1365 TACACGGACACCGGTTCTTAGCTGAGTACCTTCTCTAGCACTCCAGAGACCCCGTCCCA 1424
Db 1325 TACACGGACACCGGTTCTTAGCTGAGTACCTTCTCTAGCACTCCAGTGAACCATGCGCG 1384

QY 1425 GGGATGTTCTATGTGAAGACTGGACGGTGCATCCGAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTTCACTGTCGCCCA CGGGCGGTGTATCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCCGGATTTAGTATGATGAGCGTTTACTGCGGATGCAATGCCACCCACAGTTTC 1544
Db 1445 GCCACTGCACCGACCAACAGCATGAGTCAACTGCAGTTTGCAGCGCGCGGCCACAGTTTC 1504
QY 1545 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTGAACGACTGC 1564
QY 1605 GGGACGCAAGTGCAGGAGGGCTGCAGCTGTCTCTGCTGGGAGTTTCAAGTGTTCCTAAT 1664
Db 1565 GGACAAACAGCAGCAGGAGGGGTGCAGTTGTCCGCGCCAGACCTTCAGGTGTTCCTAAT 1624
QY 1665 GGGAGGTGTCTCCCTCAGAGCAGAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCT 1724
Db 1625 GGGAGGTGTCTCCCTCAGAGCAGAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCT 1684
QY 1725 GACGAGCTTTCATGTGACAGCTGAATGTCTCTTCTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGCTTTCATGTGACAGCTGAATGTCTCTTCTGACCAAAATATACCTACCGCTGC 1744
QY 1785 CAAAATGGCTTCTGTGAGCAAGGGCAACCTGAGTGTGATGGGAAGAGAGTGTAGC 1844
Db 1745 CTCAATGGCTTCTGTGAGCAAGGGCAACCTGAGTGTGACGGGAAGGAGAGTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAATGTGACTGTGGGCTGCGATCTTTTACCAACAGCTGC 1904
Db 1805 GACGGCTCAGATGAGAAAGGACTGCGACTGTGGGCTGGGCTCAITTCAGAGACAGGCTGT 1864
QY 1905 GTGGCTTGTGTCACCAATGCGGACGAGGGGAGTGGCCCTTGGCAGGTCAGCTTCCAGGCC 1964
Db 1865 GTTGTGTGGGCAACGATGCGGATGAGGGGAGTGGCCCTTGGCAGGTAAGCTTCATGCT 1924
QY 1965 CTGGGCGAGGCCACTTGTGTGGGCTCTGCTCATCTCTCTCTGACTGGCTGTCTGCA 2024
Db 1925 CTGGGCGAGGCCACTTGTGGTGTCTTCCCTCATCTCTCTCCAACTGGCTGTCTGCT 1984
QY 2025 GCTCATTTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGATGTGACCGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCGCAGTGTGACCGGCC 2044
QY 2085 TTCTGTGCTGTCTGTCGACCAAGCAGCAAGCGAGTGCCTTCTGGGGTGCAGGAGCTCAAGCTC 2144
Db 2045 TTCTGTGCTGTCTGTCGACCAAGCAGCAAGCGAGCGCCCTTGGGGTGCAGGAGCGCAGCTC 2104
QY 2145 AAACGTATCATCACCACTTCTTCAATGATTTTCACTTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGGCGATCATCTCCCACTTCTTCAATGACTTCACTTTCGACTATGACATCGCCTG 2164
QY 2205 CTGGAGCTGGAGAGTTCGGTGTGAGTACAGCAGCGCTGTGGCGCCCATCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAGTTCGGTGTGAGTACAGCAGCGCTGTGGCGCCCATCTGCTGCTGAT 2224
QY 2265 GCTACCCATGCTTCTTCTGTCGCAAGGCCCATCTGGGTTCAGAGCTGGGGGCAACAAAA 2324
Db 2225 GCTTCCCATGCTTCTTCTGTCGCAAGGCCCATCTGGGTTCAGAGCTGGGGGCAACAAAA 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTGCGAGAGGGTGTGATCCGTGTCTCATCAACAGACC 2384
Db 2285 TATGGAGGACTGCGCGCTGATCTTCAAAAGGGTGTGATCCCGCTCATCAACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGCGCAGAGATCAACCCCAAGATGTGTGTGGGTCTTCTC 2444
Db 2345 ACCTGTGAGGACCTCATGCGCGCAGAGATCAACCCCAAGATGTGTGTGGGTCTTCTC 2404
QY 2445 AGTGGGGGTGTGACTCTCTGCGAGGGTGTCTGTGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGCGTGGACTCTCTGCGAGGGTGTATTCGCGGGGACCCCTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGAATGTTCCAGGGCTGTGTGTGAGCTGGGGGTGAGGCTGGCTCAGAGNAC 2564

Db 2465 GATGGCGGATCTTCCAGGCGGTGTGGTGGAGAGCGGTGGGCTCAGAGGAAC 2524
QY 2565 AAGCCAGGCGGTGTACACAGGCTCCCTGTAGTTCGGGAGTGAATCAAGAGCACACTGG 2624
Db 2525 AAGCCAGGCGGTGTACACAGGCTCCCTGTAGTTCGGGAGTGAATCAAGAGAACACTGG 2584
QY 2625 GPATAGCAGCTGAGCAGACAGCCGACCAACACACACACAGGATGCCCGGACATGCACA 2684
Db 2585 GTATAGGGGCGGG--GCCACCAAAATGTGTACACCTGGCGGGCCACCCATCGTCCACC 2641
QY 2685 CTTGATACAGAGAGGAACACTGACGACATTTATGTTGGCTCCCTCCCTCCCTCCCTCC 2744
Db 2642 CCAGTGTGCACG-CCTGAGGCTGGAGACTGACCGCTGACTGCACCGCGCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAATGCACTCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735

RESULT 7

US-10-156-214A-1
; Sequence 1, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Slev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic US
; TITLE OF INVENTION: "Therof
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156, 214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSp1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H
; DATABASE ACCESSION NUMBER: GenBank AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: US Pat 5972616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-156-214A-1

Query Match 60.6%; Score 1893.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAACACCATGGTAGCAATCGGGCCGCAAGCCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGGAGGGGGCCCGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTTGGAGG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCGACGAGAAAGTGAATGGCTTTGGAGGA 124
QY 165 GGTGTGAGTTCCTGCTCGGCAATGCAAGAAAGTGGAGAGCGAGGCCCGCCAGGCGC 224
Db 125 GCGGTGGAGTTCTGCTGCACTCAACACGTCAAGAAAGGTGAAAGAGCATGGCCCGGGGCGC 184
QY 225 TGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 284

Db 185 TGGGTGCTGCTGCGACGCGTGTGATCGGCGCTCTCTCTTGGTCTTGTCTGGGATCGGCTTC 244
QY 285 CTGGTGTGCGCACTTCCATTTATCGGAATGTGGGGTTCAAAAGTCTTCAATGCGCATCTG 344
Db 245 CTGGTGTGCGCACTTTCAGTACCGGACGCTGCTGTCCAGAGGTCTTCAATGCTACATG 304
QY 345 AGGATACAAATGAGATCTTCTTGGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATACAAATGAGATTTTGTGGATGCTTACGAGAACTCCAACTCCACCTGAGTTTGT 364
QY 405 AGCTGGCGAGCAGGCTGAAGAGGCGCTGAAGTCTGTGTACAAATGAAGTCCCTGCTCTG 464
Db 365 AGCTGGCGAGCAGGCTGAAGAGCAGCGCTGAAGTCTGTACAGCGGAGTCCCATTTCTG 424
QY 465 GGTCTCTACCAAGAGTCTGCTGTAACTGCTTTCAGTGAAGGCGAGTGTCAATGCCCTAC 524
Db 425 GCGCTTACCAAGAGTCTGCTGTGAGCGCTTTCAGGAGGCGAGGCTCATGCCCTAC 484
QY 525 TACTGCTCAGAGTTTACGATCCCCCACACCTGGCAGAGAGGTTGATCGGCCCAATGGCT 584
Db 485 TACTGCTCAGAGTTTACGATCCCCCACACCTGGTGGAGAGGCGCGCTCATGGCC 544
QY 585 GTGGAGCGAGTTTGAACATTTGCCACCCCGAGCAGCGGCACTGAAATCTTCTGCTGCTACA 644
Db 545 GAGGAGCGGTAGTCAATGCTGCGCCCGCGGCGCGCTCTCTGAAAGTCTCTTGTGCTACC 604
QY 645 TCTGTGTGGCGCTTCCCCCATTTGACCCCAAGATGCTGCAGAGGACTCAGACAAACAGCTGC 704
Db 605 TCAGTGTGTGGCTTTTCCCAACGACTTCCAAACAGTACAGAGGAGCCAGGAGCAACAGCTGC 664
QY 705 AGTTTGTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 764
Db 665 AGTTTGTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 724
QY 765 AACAGTCTTACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 824
Db 725 GACAGCGCTTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 784
QY 825 GTGCTGAGCTTACCTTCCGAGCTTTGATGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCT 884
Db 785 GTGCTGAGCTTACCTTCCGAGCTTTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 844
QY 885 CTGGTACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 944
Db 845 CTGGTACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 904
QY 945 GGCACCTTCTCACCTCTCTCAACCTGACTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTACCTTCTCTCTCAACCTGACTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
QY 1005 AGCTGATTAACCAATPACTGACCGCGACATCTGCTTTTGGAGGCCACTTTTCTTCCAGCTG 1064
Db 965 ACCTGATTAACCAACACTGAGCGCGCATCTCGGCTTTTGGAGGCCACTTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGAGCTGTGGCGGCTTTTGTGATGACACCCAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGAGCTGTGGAGCGCGCTTACGTAAGCCCGAGGGGACATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCGCACTACCGCCCAACATCAACTGACATGGAATATCAAGGTGCGCCAAAC 1184
Db 1085 TACTATCCAGGCGCACTACCGCCCAACATCAACTGACATGGAATATCAAGGTGCGCCAAAC 1144
QY 1185 AACCGGAACGTGAAGGTGCTTCAAACTTTCTATCTGTTGGAGCCCAAGGTACCAAGT 1244
Db 1145 AACAGAGCTGAAGGTGAGCTTCAAACTTTCTATCTGTTGGAGCCCAAGGTACCAAGT 1204
QY 1245 GGCTCTGCAACAGGACTATGTGAGATCAACGGGAGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTTCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264
QY 1305 CAGTTTGTGTTGAGCAGCAACAGCAGAGCAAGATTAACGTCCACTTCTCTCTCTCTCTCTCT 1364
Db 1265 CAGTTTGTGTTGAGCAGCAACAGCAGAGCAAGATTAACGTCTCTCTCTCTCTCTCTCTCT 1324

```
QY 1365 TACACGACACCGGTTCTTAGCTGAGTACCTTCTCTACGACTCCAAACGACCCCGTCCCA 1424
Db 1325 TACACGACACCGGTTCTTAGCTGAATACCTTCTCTACGACTCCAGTGAACCCATGCGCG 1384
QY 1425 GGGATGTTCAATGTCAGACTGGACGCTGCATCCGAAGGAACTCGCTGCGACGCGTGG 1484
Db 1395 GGGAGTTTCACTGTCGCGACGCGGCGGTGTATCCGGAAGGAGCTCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTCCCGGATATATAGTATGAGCGGTCTACTCCGATGCAATGCCACCCACAGTTTC 1544
Db 1445 GCGGACTCCACCGACACACAGGATGAGCTCAACTGCAGTTGGACGCGGCCACCACTTC 1504
QY 1545 ACCTGCAAAACACAGTTCTGACAGCCCTCTTCTGGGTCTGTGACAGTGTCAAACAGTGT 1604
Db 1505 ACCTGCAAAACACAGTTCTGCAAGCCCTCTTCTGAGCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
QY 1605 GGGGACGGAAGTACGAGAGGGCTGACGCTGTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGAGACACAGCAGCAGAGCGGGTGCAGTTGTCCGGCCGACCTTCAGGTGTTCGAAT 1624
QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGAAAGCAACTGTGAGATGGGTCT 1724
Db 1625 GGGAGTGTCTCTGAAAGCCAGCAGTGTCAATGGGAAGGACGACTGTGGGGAAGGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1784
Db 1685 GACGAGGCTCTCTGCCCCAAGGTGAACGTCTGCTTGTACCAACACACCTTACCGCTGC 1744
QY 1785 CAAATGCGCTCTGTCTGACGAAAGGCAACCTCGAGTGTGATGGGAAGCAGCAGTGTAGC 1844
Db 1745 CTCATGGCTCTGCTTGACGAAAGGCAACCTCGAGTGTGACGGAAGGAGGACTGTAGC 1804
QY 1845 GATGCTCCGATGAGAAACACTGACTGTGGGTGCGATCTTTTACAAACAGGCTCGC 1904
Db 1805 GACGCTCAGATGAGAAAGCTGCGACTGTGGGTGCGCTCATTCACGAGACAGGCTCGT 1864
QY 1905 GTGGTGTGGCAAGAACGCGACAGCGGCGAGTGGCCCTGGCAGTGTGAGCTCCAGCC 1964
Db 1865 GTTGTGTGGGCGACGATGCGGATGAGGCGAGTGGCCCTGGCAGTGTGAGCTCCAGCT 1924
QY 1965 CTGGCCACGCGGCACTGTGTGGGCGCTCGCTCATCTCTCTGACTGTGCTCTCTGCA 2024
Db 1925 CTGGCCACGCGGCACTGTGGGTGCTCTCTCTCATCTCTCCCACTGCTGTCTCTGCC 1984
QY 2025 GCTCATCTCTTACGATGACAAAATTTCAAGTACTCAGACTACACGATGTGACGCGCC 2084
Db 1985 GCACACTGCTATCATGATGACAGAGGATTCAGGTACTCAGACCCCAAGTGTGACGCGCC 2044
QY 2085 TTCTGGGTCTGCTGACGACGAGCAGCGAGTGGCTCTGGGTGCGAGGAGTGAAGTCT 2144
Db 2045 TTCTGGGTCTGCGACGACGAGCAGCGAGCGCCCTGGGTGCGAGGAGTGAAGTCT 2104
QY 2145 AAACGTATCATCACCAACCTTCTTCAATGATTTTCACTTTCGACTATGACATCGCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTTCACTTTCGACTATGACATCGCTTG 2164
QY 2205 CTGGAGCTGGAGAACTGGTGAAGTACAGACCGTCTGTGGGCGCCCATCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGCTCCATGGTGGGCGCCCATCTGCTGCGGAC 2224
QY 2265 GCTACCCATGCTTCCCTGCTGGCAAGGCCATCTGGTTCAGAGCTGGGGGACACAAAA 2324
Db 2225 GCTTCCCATGCTTCTCCCTGCGGCAAGGCCATCTGGGTACGGGTGCGGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTCTGCAAGAGGTGAGATCCGTGTTCATCAACACAGACC 2384
Db 2285 TATGAGGACACTGGCGCTGATCTGCAAAAGGTGAGATCCGTGTTCATCAACACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGGAGCAGATACCCCAAGAACTGGAGAACGAGGCGCCAGCGC 2444
Db 2345 ACCTGTGAGAACCTCTCTGCGGAGCAGATACCGCGCGCATGATGTGGTGGGCTTCTCT 2404
```

```
QY 2445 AGTGGGCTGTGACTCTCTGCCAGGCTGACTCTGTGTGCCCCCTTGTCAAGCGCGAGAAA 2504
Db 2405 AGCGGCGCGTGGACTCTCTGCCAGGCTGATTTCCGCGGAGACCTCTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGGCGAATGTTTCCAGGCTGTGTGTGAGCTGGGGTGAAGGCTGGCTTCAGAGAAC 2564
Db 2465 GATGGGCGGATCTTTCCAGGCGCGGTGTGTGAGCTGGGGAGACGGCTGGCTTCAGAGAAC 2524
QY 2565 AAGCAGGCGGTGTACAAAGGCTCCCTGTGTGTGAGTGTGATCAAAAGACACACTGGG 2624
Db 2525 AAGCAGGCGGTGTACAAAGGCTCCCTGTGTGTGAGTGTGATCAAAAGACACACTGGG 2584
QY 2625 GTATAGCAGTGTGACAGACAGCAGCAGCAACAAACACCCACAGGATGCCACATGCA 2684
Db 2585 GTATAGGCGCGCGG---GCCACCCAAATGTGTACACTGTGCGGGGCCACCATCGTCCACC 2641
QY 2685 CCTGATACAGGAGAGGAAACACTGACGACATTTATGTGTGGCTTCCCCCCCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGAGGCTGGAGCTGGACCGCTGACTGCACACAGCGCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAATGCATCTTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGCGCTCCAAAT 2735
```

RESULT 8

```
US-10-190-030B-1
; Sequence 1, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190,030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSp1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-190-030B-1
```

```
Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCTGGGGTACCATGGGAGCGGATCGGGCGCGCAAGCGCGGAGGGGCGCGAAG 64
QY 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 124
QY 165 GGTGTGAGTTCCTGCTCGGAAACAATGCCAAGAACTGGAGAACGAGGCGCCAGCGCGC 224
Db 125 GGCGTGGAGTTCCTGCTCGGAGTCAACAAGTCAAGAGGTGGAAAGCATGTGGCGCGCGCGC 184
```


QY	225	TGGGTGGTGTGGTGGCAGTGTCTGTTT	CAGCTTCTCTTGTCTCTCCCTCAATGCGTGTG	284
Db	185	TGGGTGTGTGTGGCAGCGGTGCTCAT	CGGCTCTCTTGTGTTCTGTCTGGGATCGCGTTC	244
QY	285	CTGTGTGGCACTTCCATTATCGGAAT	TGCGGGTTCAAAAGTCTTCAATGGCCATCTG	344
Db	245	CTGTGTGGCATTTTGAGTACCGGGA	GTGGGTGTCAGAAAGTCTTCAATGGCTACATG	304
QY	345	AGGATCAAAATGAGATCTTTTCTGG	ATGCGTATGAGAACTCCACTCCACAGAGTTTATC	404
Db	305	AGGATCAAAATGAGATTTTGTGATG	TCCTACGAGAACTCCAACTCCACTGAGTTGTA	364
QY	405	AGCCTGGCAGCAGAGTGAAGGAGCG	GTGAAGTGTGATCAATGAAGTCCCTGTCTGT	464
Db	365	AGCCTGGCAGCAAGAGTGAAGGAC	CGCGTGAAGTGTGTACAGCGGAGTCCCATTC	424
QY	465	GGTCCCTTACCACAAGAAGTGGCGT	GTAACTGCTTCACTGAGGCGCATGTCGCGCTAC	524
Db	425	GGCCCTTACCACAAGAGTGGCTGT	GAGCGCTTACGAGGCGCAGCTCATCGCTTAC	484
QY	525	TACTGGTCAGAGTTTCAAGTCCCTCC	ACACCTGTGCAGAAAGGTTGATCGCGCCATGGCT	584
Db	485	TACTGTGTGAGTTTCAAGTCCCGCAG	CACCTGTGTGAGGAGGCGGAGCGCTCATGCGC	544
QY	585	GTGAGCGAGTTGTAAATTTGCCAC	CCGAGCAGCGGCACTGAAATCTCTGTGTGTAA	644
Db	545	GAGGAGCGGTAGTTCATGCTGCC	CCGCGGCGCTCCCTGAAGTCTTTGTGTGTCACC	604
QY	645	TCTGTGTGGCTTCCCATTTGACCC	CAGAAATGCTGCAGAGGACTCAGGACACACAGTGC	704
Db	605	TCAGTGTGTGCTTTCCTCCACG	GAATCCAAACAGTACAGAGGACCCAGGACACACAGTGC	664
QY	705	AGTTTGTCCCTGCAATGCCATGGT	GTGCAGCAGTGCACGCTTCACTACCCCTGGCTTCCCC	764
Db	665	AGCTTTGGCTGTACGCGCGCGGT	GTGTGAGTGTATGCTTCCACACGCGCGCTTCCCT	724
QY	765	AACAGTCCCTACCGGGCGATGCC	CGTGCAGTGGGTCTCTGCGGGGAGCGCGACTCT	824
Db	725	GACAGCCCCCTACCCCGCTCAT	GCCGCTGTCAGTGGGCGCTTGGCGGGGACGCGGACTCA	784
QY	825	GTGCTGAGCCTCACCTTCCGA	GCTTTGATGTGCTCCCTGTGATGAGCATGGCAGTGAC	884
Db	785	GTGCTGAGCCTCACCTTCG	CAGCTTTTGACCTTGGCTCTGGCAGGACGCGGAGCGAC	844
QY	885	CTGCTCACCGTGTATGATAGCT	GAGCCCCATGGAAACCCACGCTGTGTGGTGGCGCTGTGT	944
Db	845	CTGTGAGCGGTATCAACACCT	GAGCCCCCATGGAGCCCCACGCGCTGTGTGTGTGT	904
QY	945	GGCACCTTTCACCTCTCTCA	AACTGATTTCTCTCTCCAGAAAGCTTCTCTCTCCAGAA	1004
Db	905	GGCACCTTACCTCTCTCTCA	AACTGATTTCTCTCTCCAGAAAGCTTCTCTCTCTCCAGCTATC	964
QY	1005	ACGCTGATACCAATACTGAC	CGCGAGATCTTGGCTTTGAGGCGCATTTTCTTCCAGCTG	1064
Db	965	ACCTGATAACCAACACTG	AGCGCGCATCTTGGCTTTGAGGCGCATTTTCTTCCAGCTG	1024
QY	1065	CCCAAGATGAGCAGCTGTG	CGGCTTTTTCAGTGCACCAACGAGGACATTTAGCAGCCCC	1124
Db	1025	CCTAGGATGAGCAGCTGTG	GAGCGGCTTACGTAAACCCACGAGGACATTTCAACAGCCCC	1084
QY	1125	TACTATCCAGGCCACTACCG	CCCAACATCAACTGACATGGAATATCAAGGTGCCAAC	1184
Db	1085	TACTATCCAGGCCACTACCC	ACCACCAATTTGACTGTCACATGGAACATTTAGGTGCCAAC	1144
QY	1185	AACCGGAAGTGAAGTGGCTT	CAAACTCTTCTGTGTGAGACCCCAAGTACCAAGT	1244
Db	1145	AACCAGCATGTGAAGTGT	GAGCTTCAAAATCTTCTACCTGTGGAGCGCGCTGCTGGG	1204
QY	1245	GGCTCTCTGCACCAAGGACT	TGTGCAGATCAACGGGAGAGAGTACTCGGTGTGAGAGTCC	1304
Db	1205	GGCACCTTCCCCAAGGACT	CTGTGGAGATCAATGGGGAGAAATACTCGGAGAGAGGTCC	1264
QY	1305	CAGTTTGTGTGAGCAGCA	CAGCAGCAAGATTAAGTCCACTTCCATTTCTGATCACTCG	1364

[illegible]

Db 2345 ACCTGCGAGAACCTCTCTCCGACGACATCACGCCGCGCATGATGTGGTGGCTTCCTC 2404
Qy 2445 AGTGGGGTGTGGACTCTCTGCGAGGTGACTCTCTGCTGCGCCCTTGTCAAGCGCGAGAA 2504
Db 2405 AGCGCGCGTGGACTCTCTGCGAGGTGATTCGGGGGACCCCTGTTCAGGCTGGAGCG 2464
Qy 2505 GATGGCGGAATGTTCCAGGCTGGTGTGGTGAAGCTGGGGTGAAGCTGGCTTCAGAGAAC 2564
Db 2465 GATGGCGGAATCTTCCAGGCGGGTGTGGTGAAGCTGGGGAGAGCGCTGGCTTCAGAGAAC 2524
Qy 2565 AAGCCAGCGGTGTACACAGGCTCTCTGTAGTTGGGAGCTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGCGGTGTACACAGGCTCTCTGTAGTTGGGAGCTGGATCAAGAGCACTGGG 2584
Qy 2625 GTATAGCAGCATGACAGACGCGACACAAACACACACAGGATGCGCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCAATGTGTACCTGCGGGGCCACCCATCTGCCACC 2641
Qy 2685 CCTGGATACAGGAGGAGAACACTGACGACATTTATGTGTGGCTTCCCGCCCCCAACACA 2744
Db 2642 CCAGTGTGCAAG-CTGCGAGCTGGAGACTGGACCGCTGACTGCACACAGCGCCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAATGCATCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 9
US-10-190-030B-3
; Sequence 3, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190.030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSpl
US-10-190-030B-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCCAAACCATGGTAGCAATCGGGCGCGCAAGCGCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCGCGCAAGGGCGAGGGGGCCCGAAG 64
Qy 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGACATGAATGGCTTTGAGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGACATGAATGGCTTTGAGAG 124
Qy 165 GGTGTGAGTTCTCTGCTCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGGCCCCAGCGC 224
Db 125 GCGGTGAGTTCTCTGCTCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGGCCCCAGCGC 184
Qy 225 TGGGTGTGCTGGTGGCAAGTCTGTTCAGCTTCTCTTGTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGTGCTGGTGGCAAGTCTGTTCAGCTTCTCTTGTCTCTCTCTCTCTCTCTCTCT 244
Qy 285 CTGTGTGGCACTTCCATATTCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db 245 CTGTGTGGCACTTTCAGTACCGGACGTCGCTGCCAGAGGTCTTCAATGGCTACATG 304
Qy 345 AGGATCAAAATGAGATCTTCTGGAATGCTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGGATGCTTACGAGAACTCCAACTCCATGATTTGTA 364
Qy 405 AGCTGGCCAGCAGGTTGAAGGAGGCGCTCAAGCTGCTGTACAAATGAAGTCCCTGTCTG 464
Db 365 AGCTGGCCAGCAGGTTGAAGGAGGCGCTGAGCTGCTGTACAGCGAGTCCCATTCCTG 424
Qy 465 GGTCTCTTACCAAGAAAGTCCGCTGTAACTGCTTCACTGAGGCGAGTGTCAATGCCCTAC 524
Db 425 GGCCTCTTACCAAGAGTCCGCTGTGACGCGCTTTCAGCGAGGCGCAGCTCATCGCTAC 484
Qy 525 TACTGTGTCAGAGTTTACAGATCCCGCCACACTCCGCGAGAGGTTGATCGGCGCATGGCT 584
Db 485 TACTGTGCTGAGTTTACAGATCCCGCCACACTCCGCGAGGCGCAGCTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACTTGGCCACCCGAGACGCGGCACTGAAATCTTCTGTGCTAAACA 644
Db 545 GAGGAGCGCTGAGTCTATGCTGCTGCGCGCGCTCCCTGAGTCTCTTGTGGTCAACC 604
Qy 645 TCTGTGTGCGCTTCCCATTTGACCCCGAGAAATGCTGCAGAGACTCAGGACCAAGCTGC 704
Db 605 TCAGTGTGTGCTTTCCTCCACGCACTCCAAAACAGTACAGAGGACCCAGAGCAACAGCTGC 664
Qy 705 AGTTTGTGCTGATGCCATGCTGTCAGCAGTGCACGCTTCACTACCCCTGGCTTCCCT 764
Db 665 AGCTTTGGCTGTCAGCGCCCGGCTGTGGAGCTGATGCTTTCACACGCGCGCTTCCCT 724
Qy 765 AACAGTCCCTTACCCGCGCATGCCGCTGTCAGTGGTCTCTGGGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCGCTCATGCCGCTGTCAGTGGGCTGCGGGGGAGCGCGACTCA 784
Qy 825 GTGTGAGCTTCACTTCCGAGCTTGTGATGCTGCTCCCTGTGATGAGCATGAGCAGTGAC 884
Db 785 GTGTGAGCTTCACTTCCGAGCTTGTGATGCTGCTGCTGCTGCGAGCGCGGAGCGAC 844
Qy 885 CTGTGTCAGCTGTATGATAGCTTGAAGCCCATGAGCAACCCACGCTGTGTGGCTGTGT 944
Db 845 CTGTGTCAGCTGTATGATAGCTTGAAGCCCATGAGCAACCCATGAGCGCCCATGCTGTGT 904
Qy 945 GGCACCTTCTCACCTCTCTACACCTGACTTCTCTCTCCAGCAAGCTTCTCTTCTGTC 1004
Db 905 GGCACCTTCTCACCTCTCTACACCTGACTTCTCACTCTCTCCAGCAAGCTTCTCTCATC 964
Qy 1005 ACCTGATTAACCAATGACTGACCGCGGACATCTTGGCTTTGAGGCGACTTCTTCCAGCTG 1064
Db 965 ACCTGATTAACCAATGACTGACCGCGGACATCTTGGCTTTGAGGCGACTTCTTCCAGCTG 1024
Qy 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGTGAGTGACACCCCAAGGACATTTAGCAGCCCC 1124
Db 1025 CTTAGATGAGCAGCTGTGGAGCGCTTACGTAAAGCCACGAGGACATTTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCGCACTACCGCCCAACATCAATGACATGCAATGCAATGCAATGCAATG 1184
Db 1085 TACTATCCAGGCGCACTACCGCCCAACATTTGATGCTGACATGCAATGCAATGCAATG 1144
Qy 1185 AACGGAGCTGAGTGGCGCTTCAAACTTCTTATCTGTGTGAGACCCCAAGTACCAAGTG 1244
Db 1145 AACGAGCATGAGTGGAGTGGCTTCAAACTTCTTATCTGTGTGAGACCCCGGCTGTGG 1204
Qy 1245 GGTCTCTGACCAAGGACTATGAGAGATCAAGGGGAGAACTTCTGCGTGTGAGAGTCC 1304
Db 1205 GGCACCTGCGCCAGGACTAGTGGAGATCAATGGGAGAAATACTGCGGAGAGAGTCC 1264
Qy 1305 CAGTTGTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTCTGATCACTG 1364
Db 1265 CAGTTGTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCACTCAGATCACTG 1324
Qy 1365 TACAGCAGCAGCGGTTCTTAGTGTGATGATCTCTCTTACGACTTCCCAAGCAGCCGCGCA 1424

Db 1325 TACACGACACCGGCTTCTTAGCTGAATACCTCTCTCTAGCTCCAGTCCAGCCATGCCCG 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGACCGGTGCATCCGAAAGGAACTGGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTTCACTGCGCAGCGGGCGGTGTATCCGGAAGAGCTGGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCCGATATATAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1544
Db 1445 GCGCAGTGCACCGACACAGCGATGAGTCAATCTGAGTGTGGACCGCCGCGCACAGTTC 1504
QY 1545 ACGTCAAAAAACAGTTCTGCAAGCCCTCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1604
Db 1505 ACGTCAAGAAACAGTTCTGCAAGCCCTCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1564
QY 1605 GGGACGGAAGTACAGGAGGCGCTGACGTGTCTGCTGGAGTTTCAAGTGTTCAT 1664
Db 1565 GGAGACAACAGCGACGAGCGGGTSCAGTGTGCGGCCACAGCTTCAGTGTTCAT 1624
QY 1665 GGGAGTGTCTCCCTCAGAGCCGAGAGTGTATGGAAGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAGTGTCTCTCGAAAGCCAGCAGTGCATGGAAGAGGACCTGTGGGACGGGTCC 1684
QY 1725 GACGAGCTTCATGTGACAGCGTGAATGCTCTCTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGCTCTCTGCCCCAAGGTGAACGCTGCTACTTGTACCAAAACACCTACCGCTGC 1744
QY 1785 CAAATGGCTCTGTCTGACCAAGGCGCAACCTGAGTGTGATGGAAGACGAGCTGTAGC 1844
Db 1745 CTCAATGGCTCTGTCTGACCAAGGCGCAACCTGAGTGTGACGGAAGGAGCTGTAGC 1804
QY 1845 GATGCTCCGATGAGNAAACTGTGACTGTGGCTGGATCTCTTACCAAAACAGGCTGCG 1904
Db 1805 GACGGCTCAGATGAGAAGACTGCGACTGTGGCTGCGGTCTATTCACGAGACAGGCTCGT 1864
QY 1905 GTGTTGTGTGGCAAGTACGACGAGGCGAGTGGCTGCGCTGAGTGTGAGTGTGAGTGTGAG 1964
Db 1865 GTTGTGTGGGCAACGATCGGATGAGGCGAGTGGCTGCGCTGAGTGTGAGTGTGAGTGTGAG 1924
QY 1965 CTGGGCGAGGCGACATGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGTCTGTCA 2024
Db 1925 CTGGGCGAGGCGACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGCTGTCTGTGC 1984
QY 2025 GCTCATGCTTTCAGGATGACAAAATTTCAAGTACTGAGCTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGTACATGATACAGAGGATTCAGGTACTCAGACCCACGCGAGTGGACGGCC 2044
QY 2085 TTCTCTGGGTCTGTGACACAGAGCAAGCGAGTGTCTGGGTGCGAGGCTGAAGCTC 2144
Db 2045 TTCTCTGGGTCTGTGACACAGAGCCAGCGAGCGCCCTGGGTGCGAGGCGGAGCTC 2104
QY 2145 AAAAGTATCATCACCGACCCCTTCTTCAATGATTTTCACTTGCATGATGACATGCGCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTGCATGATGACATGCGCTG 2164
QY 2205 CTGGAGCTGGAGAGTCTGGTGGAGTACAGACGCTGCTGGCGCCCTATCTGCTGCTGTAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGATACAGTCCATGCTGGCGCCCTATCTGCTGCGGAC 2224
QY 2265 GCTACCCATGTCTTCTCTGCGCAGGCGCACTGGGTACAGAGTGGGGGCACACAAA 2324
Db 2225 GCTCCCATGTCTTCTCTGCGCAGGCGCACTGGGTACAGGCTGGGGGCACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGTATCTCTGCAAGAGGGGTAGATCCGTGTCTATCAACAGACC 2384
Db 2285 TATGGAGGCACTGGGCGCTGTATCTCTGCAAAAGGGGTAGATCCGCGTCTATCAACAGACC 2344
QY 2385 ACCTGTGAGGACTCATGCTGGCAGGAGATCAACCGAATGATGTGTGGTTCCTC 2444
Db 2345 ACCTGGAGAACTCTCTGCGCAGCAGATCAACCGCGCATGATGTGGGTGGGTTCCTC 2404
QY 2445 AGTGGGGGTGTGGAATCTCTGCGCAGGAGTGTGCTGTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGCTGGAATCTCTGCGCAGGAGTGTGCTGGGAGACCCCTGTCTCAGCGGTGAGCGG 2464

QY 2505 GATGGCGGAATTTTCCAGGCTGCTGTGTGTAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATTTTCCAGGCGGTGTGTGTGTAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGGCGTGTACACAGGCTCCCTGTGTGTGTAGTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 2525 AAGCCAGGCGTGTACACAGGCTCCCTGTGTGTGTGTAGTTCGGGACTGGATCAAGAGAACTGGG 2584
QY 2625 GTATAGCAGCTGACAGAGCGACAGCGACCAACCAACACAGGATGCCCGACATGCACA 2684
Db 2585 GTATAGGCGCGCGG---GCCACCCAAATGTGTACACCTGCGGGGCGCACCATGTGTCACC 2641
QY 2685 CTTGTATACAGAGAGGAACACTGACGACATTTATGCTGTGGCTCCCTCCCTCCCTCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGCGAGGCTGGAGACTGACCGCTGACTGCACCGACCGGCC-CCAGA 2699
QY 2745 ACCGAGCTGTGAACCTGATCCTTATAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

RESULT 10
US-10-267-219-1
; Sequence 1, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Yeh, Jiunn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2;
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; PRIORITY FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-267-219-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCCATGGGTAGCAATCGGGCGCCGCAAGCGCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCGCCGCAAGGGCGAGGGGGCGCCGAG 64
QY 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGCTAGAGAACTGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGCTAGAGAACTGAATGGCTTTGAGGAG 124
QY 165 GGTGTGAGTTCCTGCGTGGCAACAATGCCAAGAAAGTGGAGAGCGGAGGCCCGAGCGCC 224
Db 125 GCGGTGAGTTCCTGCGCAGTCAACACGCTCAAGAGGTGGAAAGAGCATGGCCCGGGCGC 184
QY 225 TGGGTGGTGTGGTGGAGTGTCTGCTCAGCTTCTCTGCTCTCCCTCATGGCTGGCTTG 284

185 TGGTGGTGTGGCAGCGGTGCTGATCGGCTCCTCTTGGTCTTGGTGGGATCGGCTTC 244
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352

QY 2445 AGTGGGGTGTGGACTCCTGCCAGGGTGAATCTGTGGGCCCCCTGTGCAAGCGCGAGAAA 2504
Db 2405 AGCGCGCGGTGGACTCCTGCCAGGGTGAATCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
QY 2505 GATGGCGGAATGTTCCAGGCTGGTGTGGTGAAGTGGGCTGAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGAATGTTCCAGGCGGTGTGGTGAAGTGGGCTGAGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGG 2624
Db 2525 AAGCCAGCGGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGAGAACTGGG 2584
QY 2625 GTATAGCAGCTGAGACAGACAGCCACCAACACACACAGGATGCCGACATGCA 2684
Db 2585 GTATAGGCGCGGG---GCCACCAAAATGTGTACACCTGCGGGGCGACCCATCGTCCACC 2641
QY 2685 CCTGATACAGAGAGGAACACTGACGACATTTATGTGTGGCTCCCGCCCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGACGGCTGAGACTGGACCGCTGACTGCACAGGCGCC-CCAGA 2699
QY 2745 ACCGAGCTGGAATGATCCTTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 11

US-10-267-219-3
; Sequence 3, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Yeh, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSPI

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAACCATGGTAGCAATCGGGCCCGCAGGCGGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCCGAAG 64
QY 105 GACTTCGGCGCGGGACTCAAGTACAACCTCCCGGCTAGAGACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACCTCCCGGCGAGAGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGGAGTTCCTGCTCGCAACAAATCCCAAGAAAGTGGAGAGCGGCGCCAGGCGC 224
Db 125 GCGGTGGAGTTCCTGCCAGTCAACAACTCAAGAGGTGGAGAAAGCATGGCCGGGCGC 184
QY 225 TGGGTGTGCTGGTGGAGTGTCTAGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGTGCTGGCAGCGGTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGGTGTGGCACTTCATATTCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db 245 CTGGTGTGGCACTTCATATTCGGAATGTGGGGTTCAGAGAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAAATGAGATCTTTCTGGATCGGTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCACAAAATGAGATTTTGTGGATCGCTTACGAGAACTCCAACTCCACTGAGTTTGA 364
QY 405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGTGTACAAATGAATGCTCCTCTCTG 464
Db 365 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGTGTACAGCGGATGCCATTCCTCTG 424
QY 465 GGTCCCTACCAAGAAGTCTGGCTGTAACTGGCTTCAGTGGGAGGAGTGTCTATCCGCTAC 524
Db 425 GGGCCCTACCAAGAAGTCTGGCTGTAACTGGCTTCAGTGGGAGGAGTGTCTATCCGCTAC 484
QY 525 TACTGGTTCAGAGTTCAGCATCCCCCACCTGGCAGAGAGTGTGTATCGGCCATGGCT 584
Db 485 TACTGGTTCAGAGTTCAGCATCCCCCACCTGGCAGAGAGTGTGTATCGGCCATGGCT 544
QY 585 GTGAGGAGGTGTGAACATTTCCACCCCGAGCAGCGGCACTGAAATCCTTCTGTGTAAACA 644
Db 545 GAGGAGGCGGTAGTCACTGCTGCCCGCGGCGCTCCCTGAAGTCTCTTTGTGTCTACC 604
QY 645 TCTGTGGTGGCTTTCCCAATTCAGCCAGAAATGTGTGAGAGACTCAGAGACAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCAATTCAGCCAGAAATGTGTGAGAGACTCAGAGACAACAGCTGC 664
QY 705 AGTTTTCCTGCTGATGCTGCTGAGCAGTGTGAGCAGTGTGAGTGTGAGTGTGAGTGTG 764
Db 665 AGTTTTCCTGCTGATGCTGCTGAGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 724
QY 765 AACAGTCTTACCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
Db 725 GACAGCCCTTACCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784
QY 825 GTGCTGAGCCTCAGCTTCGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 884
Db 785 GTGCTGAGCCTCAGCTTCGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 844
QY 885 CTGCTCAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 944
Db 845 CTGCTCAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 904
QY 945 GGCACCTTCTCAGCTTCCTCAACCTGACCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTCTCAGCTTCCTCAACCTGACCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
QY 1005 ACGCTGATACCAATACTGACCGGCGGATCTGCTGGCTTGGAGGCGCATTTCTTCTCAG 1064
Db 965 ACCTGATACCAATACTGACCGGCGGATCTGCTGGCTTGGAGGCGCATTTCTTCTCAG 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCCAAGGGAGCATTTAGCAGCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCGGCTTACGTAAAGCCAGGGGAGCATTCACAGCCCC 1084
QY 1125 TACTATCCAGGCGGCTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGGCGGCTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1144
QY 1185 AACCGGAACGTGAGGTCGCTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1244
Db 1145 AACCGGAACGTGAGGTCGCTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1204
QY 1245 GGTCTCTGCGACAGGACTATGTGGAGTCAACCGGGGAGAGTACTCTCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCGCCAGGACTACGTGGAGTCAATGGGGAGAAATACTCGGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGGAAACAGCAGCAAGTATACGTCCACTTCCATTTCTGATCACT 1364
Db 1265 CAGTTTGTGTGAGCAGGAAACAGCAGCAAGTATACGTCCACTTCCATTTCTGATCACT 1324
QY 1365 TACAGGACACCGGTTCTTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 1424
Db 1325 TACAGGACACCGGTTCTTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 1384

QY 1425 GGGATGTTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACCTGCGCTGCGAGCGCTGG 1484
Db |||||
QY 1385 GGGCAGTTTCATGTCGCGCACGGGCGGTGTATCCGAAAGGAGCTGCGCTGTGATGGCTGG 1444
Db |||||
QY 1485 GCAGACTGCCCGGATTAATGATGATGAGCGTTACTGCGGATGCAATGCGACCCACCGATTC 1544
Db |||||
QY 1445 GCCGACTGCACCGGACACAGCGGATGAGCTCAACTGCACTTGCAGTTCGACGCGCGGCCACCGATTC 1504
Db |||||
QY 1545 ACCTGCAAAAACCACTTCTGCAAGCCCTCTCTGCGGCTGTGACAGTGTCAAGCACTGT 1604
Db |||||
QY 1505 ACCTGCAAGAACAGTTCTGCAAGCCCTCTCTGCGGCTGTGACAGTGTGAAAGACTGC 1564
Db |||||
QY 1605 GGGGACGGAAGTGACGAGGAGGCTGCAAGCTGTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db |||||
QY 1565 GGAGACAAACGACGACGAGCGAGGGGTGCAAGTGTGCTGGGCCGACCTTCAGGTGTTCCAAT 1624
Db |||||
QY 1665 GGGAGTGTCTCCCTCAGACGCGAGCAAGTGTAAATGGGAAGGACAACTGFGGAGATGGGTCT 1724
Db |||||
QY 1625 GGGAGTGCCTCTCGAAAGCCGACGAGTGCATGGGAAGGACGACTGTGCGGAGCGGGTCC 1684
Db |||||
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGCAACCAATATACCTACCGCTGC 1784
Db |||||
QY 1685 GACGAGGCTCTCTGCCCAAGGTGAACTGTCTGCTTGTACCAACACACCTACCGCTGC 1744
Db |||||
QY 1785 CAAATGSCCTCTGTGTGAGCAAGGCAACCTGAGTGTGTGATGGGAAGGACGACTGTAGC 1844
Db |||||
QY 1745 CTCAATGGCTCTGTGTTGAGCAAGGCAACCTGAGTGTGACGGGAAGGAGACTGTAGC 1804
Db |||||
QY 1845 GATGCTCCGATGAGAAATCTGACTGTGCGCTGCGATCTTGTACCAACACAGCTGC 1904
Db |||||
QY 1805 GACGCTCAGATGAGAGGACTGCGACTGTGGGTGCGGTCAATTCAGAGACAGGCTCGT 1864
Db |||||
QY 1905 GTGTTGTGTCGCAATATCGGACGAGGCGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db |||||
QY 1865 GTTGTGGGGGACGAGTCCGATGAGGCGAGTGGCCCTGGCAGGTAAAGCTGCATGCT 1924
Db |||||
QY 1965 CTGGGCGAGGCGCACTGTGTGGGCGCTGCTCATCTCTCTGACTGCTGCTCTGCA 2024
Db |||||
QY 1925 CTGGGCGAGGCGCACTGTGGGCGCTTCTCCCTCATCTCTCCCACTGGCTGTCTGCTGCC 1984
Db |||||
QY 2025 GCTCATGCTTTCAGGATGACAAATTTCAAGTACTCAGACTACAGATGTGACGCGCC 2084
Db |||||
QY 1985 GCACACTGCTATCATGATGACAGAGGATTCAGGTACTCAGACCCCGCAGTGTGACGCGCC 2044
Db |||||
QY 2085 TTCTGGGCTGTGTCGACGACGAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTCAAGCTC 2144
Db |||||
QY 2045 TTCTGGGCTTGCACGACGAGCGAGCGAGCGCCCTGGGGTGCAGGAGCGAGGCTC 2104
Db |||||
QY 2145 AAAGTATCATACCGACCCCTCTCTCATGATTTCACTTCGACTATGACATCGCCTTG 2204
Db |||||
QY 2105 AAGCGCATCATCTCCCAACCCCTTCTCAATGACTTCACTTCGACTATGACATCGCGCTG 2164
Db |||||
QY 2205 CTGGAGCTGGAGAAAGTGGTGGAGTACAGACCGCTGCTGGCGCCCATCTGCTCCCTGAT 2264
Db |||||
QY 2165 CTGGAGCTGGAGAAACCGGAGAGTACAGCTTCATGTGGGCGCCCATCTGCTCCCGAC 2224
Db |||||
QY 2265 GCTACCATGCTTCTCTGTGGCAAGGCCATCTGGGTTCACAGCTGCGGGGSCACACAAA 2324
Db |||||
QY 2225 GCCTCCCATGCTTCTCTCTGGCGGCAAGGCCATCTGGGTTCAGGGCTGGGGACACACCCAG 2284
Db |||||
QY 2325 GAGGAGGTACCGGAGCGCTGATCTCTGAGAGGAGGTGAGATCCGTGTCTATCAACAGACC 2384
Db |||||
QY 2285 TATGAGGCACTGGCGGCTGATCTCTGCAAAAGGTTGAGATCCGCGCTCATCAACAGACC 2344
Db |||||
QY 2385 ACCTGTGAGGACCTCATGCGCGCAGAGATCAACCCACGAAATGATGTGTGGGTTCCTC 2444
Db |||||
QY 2345 ACCTGCGAGAACCTCTGCGCGCAGAGATCAACCGCGGAGATGATGTGGGTTCCTC 2404
Db |||||
QY 2445 AGTGGGCTGTGACTCTCTGCGAGGCTGACTCTGTGTGCGCCCTTTGTCAAGCGCGAGAAA 2504
Db |||||
QY 2405 AGCGGCGCGTGGACTCTCTGCGAGGTTGATTCGCGGGGACCCCTGTCTCAGCGTGGAGCG 2464
Db |||||

QY 2505 GATGGCGAATGTTTCCAGGCTGTGTGTTGAGCTGGGTGAAGGCTGCGCTCAGAGAAC 2564
Db |||||
QY 2465 GATGGCGGATCTTTCAGGCGCGGTGTGTGAGCTGGGAGACGCGTGTGCTCAGAGAAC 2524
Db |||||
QY 2565 AAGCCAGCGTGTACACAAAGGCTCCCTGTGTAGTTCGGGACTGGATCAAAGAGACACACTGGG 2624
Db |||||
QY 2525 AAGCCAGCGTGTACACAAAGGCTCCCTGTGTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584
Db |||||
QY 2625 GTATAGCAGCATGAGACAGACAGCGGACACAAACACCCACAGGATGCCCGACATGCACA 2684
Db |||||
QY 2585 GTATAGGGGGCGGG--GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641
Db |||||
QY 2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGTGTGGCTCCCGCCCAACACA 2744
Db |||||
QY 2642 CCAGTGTGCAGC-CCTGACGCTGGAGACTGGACCGTGTACTGCACGAGGCC-CCAGA 2699
Db |||||
QY 2745 ACCGAGACTGTGAACCTCATCTTTAGGACTCAGAGT 2780
Db |||||
QY 2700 ACATACACTGTGAACCTCAATCTCCAGGCTCCAAT 2735
Db |||||

RESULT 12
US-10-302-840A-1
; Sequence 1, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Org. Edgar O.
; APPLICANT: Madison, Edwin L.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODE
; FILE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank #AF081724
; DATABASE ENTRY DATE: 2000-08-31
; US-10-302-840A-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCCGCAAAACCATGGGTAGCAATCGGGGCGCGAAGCGCGAGGGGGCTCTCAG 104
Db |||||
QY 5 GAGCGCGCTCGGGGTACCATGGGAGCGATCGGGCGCGCAAGGGCGGAGGGGCGCCGAAG 64
Db |||||
QY 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAG 164
Db |||||
QY 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGACAGAGAAAGTGAATGGCTTTGAGGAA 124
Db |||||
QY 165 GGTGTGGAGTTCCTGCTGCGAAACAATGCCAGAAAGTGGAGAGGAGGCGCCAGCGC 224
Db |||||
QY 125 GCGTGGAGTTCCTGCGCAGTCAACAGCTCAAGAGTGGAGAAAGCATGGCCGGGGGCGC 184
Db |||||
QY 225 TGGGTGTGTGTGGGAGTGTTCAGTTCCTCTTGTCTTCCCTCATGGCTGGCTTG 284
Db |||||
QY 185 TGGGTGTGTGTGGGAGCGGTGTGATCGGCTCTCTTGTGTGTGGGAGTGGGCTTC 244
Db |||||
QY 285 CTGCTGTGGCACTTCCATTTATCGAATGTGCGGTTCAAAAGTCTTCAATGGCCATCTG 344
Db |||||
QY 245 CTGCTGTGGCACTTTCAGTACCGGAGCGTGGGTGTCCAGAGGTCTTCAATGGCTACATG 304
Db |||||

QY 345 AGGATCAAAATGAGATCTTTCTGGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db AGGATCAAAATGAGATCTTTCTGGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
QY 405 AGCTGGCCAGCAGGATGAGAGAGGCTGAGCTGCTGTATGAGTGAAGTCCCTGCTGCTG 464
Db AGCTGGCCAGCAGGATGAGAGAGGCTGAGCTGCTGTATGAGTGAAGTCCCTGCTGCTG 464
QY 465 GGTCTCTACCAAGAGTGGCTGTAACTGGCTTTCAGTGGAGGAGTGTATGCTGCTTAC 524
Db GGTCTCTACCAAGAGTGGCTGTAACTGGCTTTCAGTGGAGGAGTGTATGCTGCTTAC 524
QY 525 TACTGGTCAAGATTCAGATCCCTCCACACCTGGCAGAGAGTGTATGCTGCTGCTGCTGCT 584
Db TACTGGTCAAGATTCAGATCCCTCCACACCTGGCAGAGAGTGTATGCTGCTGCTGCTGCT 584
QY 585 GTGGAGCAGATGTAACTTGCACCCGAGCAGCAGGCTGAGTGAATCTTCTGCTGCTTAC 644
Db GTGGAGCAGATGTAACTTGCACCCGAGCAGCAGGCTGAGTGAATCTTCTGCTGCTTAC 644
QY 645 TCTGTGGTGGCTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 704
Db TCTGTGGTGGCTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 704
QY 705 AGTTTGGTGGCTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 764
Db AGTTTGGTGGCTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 764
QY 765 AACAGTCTCTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 824
Db AACAGTCTCTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 824
QY 825 GTGCTGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 884
Db GTGCTGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 884
QY 885 CTGGTCAAGTGTATGATGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 944
Db CTGGTCAAGTGTATGATGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 944
QY 945 GGCACCTTCTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1004
Db GGCACCTTCTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1004
QY 1005 ACCTGATTAACCAATGATGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1064
Db ACCTGATTAACCAATGATGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1064
QY 1065 CCAAGATGAGAGTGTGGGCTTTTGGAGTGAACCCCAAGGAGCAATTTAGCAGGCTT 1124
Db CCAAGATGAGAGTGTGGGCTTTTGGAGTGAACCCCAAGGAGCAATTTAGCAGGCTT 1124
QY 1125 TACTATCCAGCCACTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1184
Db TACTATCCAGCCACTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1184
QY 1185 AACCGAAGTGAAGTGTGGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1244
Db AACCGAAGTGAAGTGTGGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1244
QY 1245 GGCTCTGCAAGAGTGTATGAGATGAGAGGAGAGTGTATGAGAGTGTATGAGAGTGTATG 1304
Db GGCTCTGCAAGAGTGTATGAGATGAGAGGAGAGTGTATGAGAGTGTATGAGAGTGTATG 1304
QY 1305 CAGTTTGTGGTGTAGCAGCAACAGCAGCAGAGTGTATGAGTGTATGAGTGTATGAGTGTATG 1364
Db CAGTTTGTGGTGTAGCAGCAACAGCAGCAGAGTGTATGAGTGTATGAGTGTATGAGTGTATG 1364
QY 1365 TACAGGACACCGGCTTCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1424
Db TACAGGACACCGGCTTCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1424
QY 1325 TACACCGACACCGGCTTCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1384
Db TACACCGACACCGGCTTCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1384

QY 1425 GGGATGTTTCATGTGCAGACTGGACGGTGCATCCGAAAGGAACTCGCTGCGACGGCTGG 1484
Db GGGATGTTTCATGTGCAGACTGGACGGTGCATCCGAAAGGAACTCGCTGCGACGGCTGG 1484
QY 1485 GCAGACTGCCGGATATAGTATGATGAGGCTTACTGCCCATGATGCAATGCCACCACTTC 1544
Db GCAGACTGCCGGATATAGTATGATGAGGCTTACTGCCCATGATGCAATGCCACCACTTC 1544
QY 1545 ACCTGCAAAACCACTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db ACCTGCAAAACCACTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
QY 1605 GGGACGGAAGTGTGAGAGGCTTGCAGCTGTCTGCTGGGAGTTCACAGTGTTCCTCAAT 1664
Db GGGACGGAAGTGTGAGAGGCTTGCAGCTGTCTGCTGGGAGTTCACAGTGTTCCTCAAT 1664
QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAGTGTATGAGAGGAACTGTGAGAGTGGGTCT 1724
Db GGGAGTGTCTCCCTCAGAGCCAGAGTGTATGAGAGGAACTGTGAGAGTGGGTCT 1724
QY 1725 GACGAGGCTTTCATGTGACAGCGTGAATGTCTCTCTTGCACCAATATACCTACCGCTGC 1784
Db GACGAGGCTTTCATGTGACAGCGTGAATGTCTCTCTTGCACCAATATACCTACCGCTGC 1784
QY 1785 CAAATGSCCTTCTGTGAGCAAGGCAACCTGAGTGTGATGAGGAACTGTGAGTGTAGC 1844
Db CAAATGSCCTTCTGTGAGCAAGGCAACCTGAGTGTGATGAGGAACTGTGAGTGTAGC 1844
QY 1845 GATGGCTCCGATGAGAAACTGTGATGAGTGGGTGTGATGAGTGTGATGAGTGTGATGAGTGT 1904
Db GATGGCTCCGATGAGAAACTGTGATGAGTGGGTGTGATGAGTGTGATGAGTGTGATGAGTGT 1904
QY 1905 GTGGTGTGTCAGCAATGCGGACGAGGCGAGTGGCTTGGCAGGCTGAGCTTCCACGCC 1964
Db GTGGTGTGTCAGCAATGCGGACGAGGCGAGTGGCTTGGCAGGCTGAGCTTCCACGCC 1964
QY 1965 GTTGTGGGCGCAGGATGCGGATGAGGCGAGTGGCTTGGCAGGCTGAGTGTGATGAGTGT 1924
Db GTTGTGGGCGCAGGATGCGGATGAGGCGAGTGGCTTGGCAGGCTGAGTGTGATGAGTGT 1924
QY 1965 GTGGCGCAGGCGCTTGTGAGGCGCTTGTGAGTGTGATGAGTGTGATGAGTGTGATGAGTGT 2024
Db GTGGCGCAGGCGCTTGTGAGGCGCTTGTGAGTGTGATGAGTGTGATGAGTGTGATGAGTGT 2024
QY 2025 GCTCATTTGCTTACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2084
Db GCTCATTTGCTTACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2084
QY 2085 TTCTGCTGCTGTGCTGAGCAAGCAGAGTGTGCTTCTGGGTGAGGAGTGTGAGTGTGAGTGT 2144
Db TTCTGCTGCTGTGCTGAGCAAGCAGAGTGTGCTTCTGGGTGAGGAGTGTGAGTGTGAGTGT 2144
QY 2145 AAAGTGTATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2204
Db AAAGTGTATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2204
QY 2205 CTGAGGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2264
Db CTGAGGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2264
QY 2265 GCTACCTTCTTCTGCTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2324
Db GCTACCTTCTTCTGCTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2324
QY 2325 GAGGAGGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2384
Db GAGGAGGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2384
QY 2385 ACCTGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2444
Db ACCTGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2444
QY 2445 AGTGGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2504
Db AGTGGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2504
QY 2405 AGCGGCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2464
Db AGCGGCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2464
QY 2505 GATGGGCGAATGTTTCCAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2564
Db GATGGGCGAATGTTTCCAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2564


```
Db 2465 GATGGCGGATCTTCCAGGCGGCTGTGGTGAAGTGGGAGACGGCTGGCTCAGAGAAC 2524
Qy 2565 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACTGGG 2584
Qy 2625 GTATAGCAGCATGACAGACAGCGACCAAAACACCCACAGGAGTCCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTGGCGGCGCCACCCCATGCTCCACC 2641
Qy 2685 CCTGGATACAGGAGGAGNACACTGACGACATTTATGCTGTGGCTCCCGCCCCCAACACA 2744
Db 2642 CCACTGTGCAGC-CCTGACGCTGGAGACTGGACCGCTGACTGCACACGAGCGCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAAGTGCATCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGCTCCAAAT 2735

RESULT 13
US-10-302-840A-3
; Sequence 3, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCOD
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSPL
US-10-302-840A-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred.No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCCAAACGATGGGTAGCAATCGGGCGCGAAGCGCGGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGCGCGAAG 64
Qy 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAG 124
Qy 165 GGTGTGAGTTCCTCGCTCGGAAACATGCCAAGAAAGTGGAGAAAGCGAGGCCCGCGCGC 224
Db 125 GGCCTGGAGTTCCTCGCCAGTCAACAGTCAAGAGGTGGAAAGCATGATGGCTTGGAGGAA 184
Qy 225 TGGGTGTGCTGTGGGAGTGTGCTGAGCTTCCTCTGCTCTCCCTCATGCGTGGCTTG 284
Db 185 TGGGTGTGCTGTGGGAGCGCTGCTGATGGCTCTCTTGGTCTTGGTGGGATCGGCTTC 244
Qy 285 CTGTGTGGCACTTCCATATATGGAATGTGGGCTTCAAAAGTCTCAATGGCCATCTG 344
Db 245 CTGTGTGGCAATTTGAGTACCGGAAGTGTGCTGCTCAGAGGTCTTCAATGGCTCATG 304
Qy 345 AGGATCAAAATGAGATCTTTCTGGATGCGGTATGAGAACTCCACTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGGATGCGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
```

```
Qy 405 AGCTGTGCAGCGAGGTGAAGGAGGCGCTCAAGCTGCTGTACAAATGAAGTCCCTGTCTCTG 464
Db 365 AGCTGTGCAGCGAGGTGAAGGAGGCGCTCAAGCTGCTGTACAGCGAGTCCCTATTCCTG 424
Qy 465 GGTCCCTTACCAAGAGTGGGTGTAACTGCTTCACTGAGGCGAGTGTATCGGCTAC 524
Db 425 GGCCTCTTACCAAGAGTGGGTGTGACGCGCTTCACTGAGGCGAGCGTCACTGCGCTAC 484
Qy 525 TACTGTGTCAAGTTCAGCATCCCGGACACCTGGCAGAAAGGTTGATCGGGCGATGGCT 584
Db 485 TACTGTGTCTGAGTTCAGCATCCCGGACACCTGGTGGAGAGCGGAGCGCGTCACTGGCC 544
Qy 585 GTGGAGCGAGTGTGTAACTATGCGCACCCCGGACGCGGCACTGAAATCTCTTGTGCTAAACA 644
Db 545 GAGGAGCGGTAGTGTATGCTGCTGCGCGGCGGCTCCCTGAAGTCTCTTGTGGTCAACC 604
Qy 645 TCTGTGTGTGCTTCCCATTTGACCCCGGCGGCTCCCTGAAGTCTCTTGTGGTCAACC 704
Db 605 TCAGTGTGTGCTTTCCTCCACGAGTCCAAAACAGTACAGAGGACCCAGGACAAACAGCTGC 664
Qy 705 AGTTTTGCTGCTGATGCCATGCTGCGAGTGTGAGCTGTGAGTGTGAGTGTGAGTGTGAGT 764
Db 665 AGTTTTGCTGCTGATGCCATGCTGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 724
Qy 765 AACAGTCTCTTACCCGCGCATGCTGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 824
Db 725 GACAGCGCTTACCCGCTCATGCTGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 784
Qy 825 GTGCTGAGCTTCACTTCCGAGCTTTGAGTGTGCTGCTGCTGCTGATGAGCATGAGGAGTGC 884
Db 785 GTGCTGAGCTTCACTTCCGAGCTTTGAGTGTGCTGCTGCTGATGAGCATGAGGAGTGC 844
Qy 885 CTGCTGACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 944
Db 845 CTGCTGACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904
Qy 945 GGCACCTTCTACCCCTCTTACAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1004
Db 905 GGCACCTTCTACCCCTCTTACAACTGATGATGATGATGATGATGATGATGATGATGATGAT 964
Qy 1005 ACCTGATTAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1064
Db 965 ACCTGATTAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1024
Qy 1065 CCAAGATGAGCAGCTGTGGCGCTTTTTCAGTGACACCCCAAGGACATTTAGCAGGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGCGCTTACGTAAGCCCAAGGACATTTCAACAGGCCCC 1084
Qy 1125 TACTATCCAGGCGCACTACCCGCGCAACATCAATGCAATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCCGCGCAACATTTGACTGCAATGGAATATCAAGGTGCCCAAC 1144
Qy 1185 AACCGGAACCTGAGGTGGCTTCAAACTTCTATCTGTTGAGCCCAACCGTACCGAGTG 1244
Db 1145 AACCGAGATGAGGTGGCTTCAAACTTCTATCTGTTGAGCCCAACCGTACCGAGTG 1204
Qy 1245 GGCTCTCTCCACCAAGGACTATGTGGAGATCAACGGGAGAAAGTACTGCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCGCCCAAGGACTACGTGGAGATCAATGGGAGGAAATACTGCGGAGAGAGTCC 1264
Qy 1305 CAGTTTGTGTGAGCGCAACAGCAAGATTTACAGTGCATTTCCATTTCTGATCACTCG 1364
Db 1265 CAGTTTGTGTGAGCGCAACAGCAAGATTTACAGTGCATTTCCATTTCTGATCACTCG 1324
Qy 1365 TACACGACACCGGCTTCTAGCTGAGTACCTCTCTAGACTCCCAACGACCGGCGCCA 1424
Db 1325 TACACGACACCGGCTTCTAGCTGAGTACCTCTCTAGACTCCCAACGACCGGCGCG 1384
Qy 1425 GGGATGTTTCATGTGCAAGACTGGACGCTGCATCCGAAAGGAACTGGCTGCGAGCGGTGG 1484
Db 1385 GGGCAGTTTACGTTGCCGACGCGGCGGTGTATCCGGAAGGAGTGTGCTGTGATGCTGG 1444
```

QY 1485 GCAGACTCCCGGATATAGTAGAGGCTTACTCCGATGCAATGCGCCACCCACAGTTC 1544
Db 1445 GCCGACTGCACCGACCAAGCGATGAGCTCACTGAGTTCGACGCGCGCCACAGTTC 1504
QY 1545 ACGTGCAAAACCAAGTCTCTGCAAGCCCTCTCTGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTGCAGAAACAAGTCTCTGCAAGCCCTCTCTGGTCTGTGACAGTGTCAACGACTGT 1564
QY 1605 GGGGACGGAAGTGAAGGAGGAGGCTGTCTGCTGGGAGTTTCAAGTGTTCAT 1664
Db 1565 GGAGACAAACAGCAGCAGGAGGAGGAGTGTCTGCGGCCAGACCTTCAGTGTTCAT 1624
QY 1665 GGGAGTGTCTCCCTCAGACGACGAGTGTATGGGAAGGACAACTGTGGAGAGTGGTCT 1724
Db 1625 GGGAGTGTCTCTCGAAAGCCAGCAGTGTGCAATGGGAAGGACGAGTGTGGGAGCGGTCT 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTGCTGCAACCAATATACCTACGCTGC 1784
Db 1685 GACGAGGCTCTGCGCCAGAGTGAAGTGTCTCTGCTGCAACCAATATACCTACGCTGC 1744
QY 1785 CAAATGGCCTCTGTCTGAGCAAGGCAACCTCTGAGTGTGATGGGAAGCAGGACTGTAGC 1844
Db 1745 CTCATGGCCTCTGTCTGAGCAAGGCAACCTCTGAGTGTGAGGAGGAGGACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAACCTGTGAGTGTGGCTGCGATCTCTTACAAACAGGCTGCG 1904
Db 1805 GACGCTCAGATGAGAGGACTGCGACTGTGGCTGCGCTCAATTCAGCAGACAGGCTCGT 1864
QY 1905 GTGGTGTGTCACAAATGCGGACGAGGCGAGTGGCCCTGCGAGGTGAGCTCCACGCC 1964
Db 1865 GTTGTGGGSCACGGATGCGGATGAGGSCGAGTGGCCCTGCGAGGTAGCCTGATGCT 1924
QY 1965 CTGGGCGAGGCCACTTGTGTGGGCGCTCGCTCATCTCTCTGAGTGGCTGTCTGCA 2024
Db 1925 CTGGGCGAGGCCACTGTGGTGTCTTCCCTCATCTCTCCCACTGCTGTCTGCC 1984
QY 2025 GCTCATTTCTTACGATGACAAATTTCAAGTACTCAGACTACAGATGTGAGCGCC 2084
Db 1985 GCACACTGTACATCGATGACAGAGGATTCAGTACTCAGACCCCGCAGCTGGACGCC 2044
QY 2085 TTCTGGTCTGTGGACGAGCAAGCGAGTGCCTCTGGGCTGCGAGGCTGAGCTC 2144
Db 2045 TTCTGGTCTGTGGACGAGCAAGCGAGTGCCTCTGGGCTGCGAGGCTGAGCTC 2104
QY 2145 AAAGTATATCAACCAACCTCTCTCAATGATTTCACTTCGACTATGACATCGCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTCTCTCAATGACTTCACTTCGACTATGACATCGCTTG 2164
QY 2205 CTGGAGCTGGAGAGTGGTGGATGACAGCAGCTGTGCGCCCATCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAGTGGTGGATGACAGCTTCCATGTTGCGGCCCATCTGCTGCGGAC 2224
QY 2265 GCTACCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCAAGGCTGGGGSCACAAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGCGGCAAGGCCATCTGGGTCAAGGCTGGGGSCACAA 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGCTGAGATCCCGTGTATCAACAGAC 2384
Db 2285 TATGGAGGCACTGGCGGCTGATCTCTGCAAAAGGCTGAGATCCCGTGTATCAACAGAC 2344
QY 2385 ACCTGTGAGGACCTCATGCGGACAGATCAACCCAGATGATGTGTGGGTTTCCCTC 2444
Db 2345 ACCTGTGAGGACCTCATGCGGACAGATCAACCCAGATGATGTGTGGGTTTCCCTC 2404
QY 2445 AGTGGGCTGTGAGTCTCTGCGAGGCTGACTCTGTGCGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGCTGTGACTCTGTGAGGCTGATCTGCGGAGACCCCTGTCTCAGGCTGAGGCG 2464
QY 2505 GATGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGCTGAAGGCTGCGCTCAGAGAAC 2564
Db 2465 GATGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGCTGAAGGCTGCGCTCAGAGAAC 2524
QY 2565 AAGCCAGCGTGTACAAAGCTCCCTGTGTGAGTTCGGGACTGTGATCAAAAGACACTGGG 2624

Db 2525 AAGCCAGCGTGTACAAAGCTCCCTGTGTTCGGGACTGTGATCAAAAGACACTGGG 2584
QY 2625 GTATAGCAGCATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2684
Db 2585 GTATAGGCGCGCGG---GCCACCAAAATGTGTACACCTGCGCGGCGCACCCATCGTCCACC 2641
QY 2685 CCTGGATACAGGAGAGGAAACACTGACGACATTTATGTGTGCTGCCCTCCCGCCCAACACA 2744
Db 2642 CCATGTGTGACG-CCTGACGCTGGAGACTGGACCGCTGACTGCACAGCGCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAAGTCTTCTTGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAAGTCTTCTTCCAGGCTCCAAAT 2735

RESULT 14
US-10-112-221A-1
; Sequence 1, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Org, Edgar O.
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9,
; FILE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSPI
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-112-221A-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCGAG 64
QY 105 GACTTCGGCGCGGAGTCAAGTACAACTCCCGGTAGAGAACATGAATGGTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGTCAAGTACAACTCCCGGTAGAGAACATGAATGGTTGAGGAG 124
QY 165 GGTGTGAGGTTCCTGCTGCGAACAATGCCAAGAGTGGAGAGCGAGGCCCGCAGCGC 224
Db 125 GCGTGGAGTTCCTGCGAACAAGTGGAGAGCGAGGCCCGCAGCGC 184
QY 225 TGGGTGCTGTGGTGGAGTGTGTTGAGTTCCTCTTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGCTGTGGTGGAGCGGTGCTGATCGGCTCTCTTGTCTTGTCTGCGGATCGGCTC 244
QY 285 CTGGTGTGGCATTTCATATTCGAGATGTGCGGTTTCAAAAGTCTTCAATGGCATTCTG 344

Db 245 CTGGTGTGGCATTTTCAGATACCGGAGCGTGGGTGTCCAGAAAGGTTCTCAATGGCTTACATG 304
QY 345 AGGATCAAAATAGATCTTTCTGATGCGTATAGAAATCTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATAGAAATTTGTGATGCGTACGAGAACTCCAATCTCACTGAGTTTGTGA 364
QY 405 AGCTGGCCAGCAGGTGAAGAGGCGGTGAAGCTGTGTAAATGAAGTCCCTGTCTCTG 464
Db 365 AGCTGGCCAGCAGGTGAAGAGGCGGTGAAGCTGTGTAAATGAAGTCCCTGTCTCTG 424
QY 465 GGTCCCTACCAAGAAGTCCGCTGTAACTGCTTCACTGAGGAGAGTGTATGCGCTTAC 524
Db 425 GCGCTTACCAAGAAGTCCGCTGTGAGCGCTTTACGCGAGGCGAGCTCATGCGCTTAC 484
QY 525 TACTGGTCAGAGTTTCAGCATCCCCCACCACCTGGCAGAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTTCAGCATCCCGCAGACCTGGTGGAGAGGCGGAGCGGTATGGCC 544
QY 585 GTGAGCGAGTTGTAAATTTGGCACCACCGAGACCGGGCACTGGAATCTTCTGTGCTAA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCGCGGCGCGCTCCCTGAAAGTCCCTTGTGTCA 604
QY 645 TCTGTGTGGGCTTCCCATTTGACCCAGAAATGCTGCAGAGAGCTTCAGGACAAAGCTGC 704
Db 605 TCAGTGTGGGCTTTCCCGCAGCTCCAAAACAGTACAGAGGAGCCCGAGGACAAAGCTGC 664
QY 705 AGTTTGGCCCTGTCATGCCCATGTTGTCAGCAGTGAACAGCTTCACTACCCCTGGCTTCCC 764
Db 665 AGCTTGGGCTTGACCGCCGCGGTGTGAGCTGATGCGCTTCAACACCGCCGGCTTCCCT 724
QY 765 AACAGTCCCTACCGCGCATGCCGCTGCAGTGGGTCTTGGGGGGGAGCGCGAGTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGSCCGCTGCAGTGGGCTTGGGCGCTTGGGGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCTTCACTTCCGAAGCTTTGATGTGCTCCCTGATGATGATGATGATGATGATGAT 884
Db 785 GTGCTGAGCTTCACTTCCGAGCTTTTGAAGCTTGTGCTTCCGAGAGCGGCGGAGCGAC 844
QY 885 GTGCTGACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 944
Db 845 CTGGTGAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904
QY 945 GGCACCTTCTACACCTCTTACAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1004
Db 905 GGCACCTTACCTCTCTTACAACTGATGATGATGATGATGATGATGATGATGATGATGAT 964
QY 1005 ACGTGTATTAACCAATACTGACCGGCGCATCTTGGCTTTTGGGCGCATTTTCTTCCAGCTG 1064
Db 965 AACTGTATTAACCAATACTGACCGGCGCATCTTGGCTTTTGGGCGCATTTTCTTCCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGGTGAACACCCAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCAAGATGAGCAGCTGTGGAGCGGCTTACGTAAAGCCAGGAGCAATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCACTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGGCACTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1144
QY 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGTGGACCCCAACAGTACCAGTG 1244
Db 1145 AACCGAGATGTAAGGTGAGCTTCAAACTCTTCTATCTGTGGACCCCAACAGTACCAGTG 1204
QY 1245 GGCTCTTGACCAAGGATATGTGGAGATCAACCGGAGAGTACTCGCGTGAAGGTCC 1304
Db 1205 GGCACTCTGCCCAAGGATACGTGGAGATCAATGGGAGAAATACTCGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAAGCAGCAGAGATTAAGTCCATTCATCTGATTCATCTG 1364
Db 1265 CAGTTTGTGTGAGCAGCAAGCAGCAGAGATTAAGTCCATTCATCTGATTCATCTG 1324
QY 1365 TACACGACACCGGCTTCTAGCTGAATACCTCTCTTACGACTCCAGACCGCGTGGCCA 1424
Db 1325 TACACGACACCGGCTTCTAGCTGAATACCTCTCTTACGACTCCAGCTCCAGTCCG 1384

QY 1425 GGGATGTTTCTGTCAAGACTGGAGCGTGTCCGAGGAACTGGCTGCGACGGCTGG 1484
Db 1385 GGGAGTTTCACTGTCGGCAGCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGTGGCTGG 1444
QY 1485 GCAGACTGCCCGGATTTAGTGTAGAGCGTTTACTTGCAGATGCAATGCGCCACCCAGTTTC 1544
Db 1445 GCGGACTGCACCGACACAGCGATGAGCTCAATGCGAGTTGCGAGCGCGGCCACAGTTTC 1504
QY 1545 ACGTGCAAAAACAGTTTCTGCAAGCCCCCTTCTTGGGTCTGTGACAGTGTCAACAGACTGT 1604
Db 1505 ACGTGCAAAAACAGTTTCTGCAAGCCCCCTTCTTGGGTCTGTGACAGTGTCAACAGACTGT 1564
QY 1605 GGGGACCGAAAGTACGAGGAGGCTGAGCTGTCTCTGCGAGTTTCAAGTTTCCCAAT 1664
Db 1565 GGAGCAACAGCGACGAGGAGGCTGAGTTGTTCGGCCCGACAGCTTCAAGTTTCCCAAT 1624
QY 1665 GGGAAAGTGTCTTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTTCCGAAAAGCCAGAGTGTCAATGGGAAGGAGAGTGTGGGACGGGTCC 1684
QY 1725 GACGAGCTTTCATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGTGC 1784
Db 1685 GACGAGCTTTCATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGTGC 1744
QY 1785 CAAAATGGCTTCTGTCTGAGCAAGGGCAACCTCTGAGTGTGATGGGAAGGACAACTGTGGAGATGGGTCT 1844
Db 1745 CTCAATGGCTTCTGTCTGAGCAAGGGCAACCTCTGAGTGTGATGGGAAGGAGAGTGTGAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
Db 1805 GADGGCTCAGATGAGAAGGACTGCGACTGTGGGTGCGGTCAATTCAGAGACAGGCTCGT 1864
QY 1905 GTGGTGTGGCGCAATGCGAGCGGAGTGGGCGGTGGCGGCTGGCAGGTGAGGCTCCAGCC 1964
Db 1865 GTGGTGTGGCGCAATGCGAGTGGGCGGTGGCGGCTGGCAGGTGAGGCTCCAGCC 1924
QY 1965 CTGGGCGAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGTCTCTGCA 2024
Db 1925 CTGGGCGAGGCGCACATCTGCGGTGTCTTCCCTCATCTCTCCCAACTGCTGTCTGCGC 1984
QY 2025 GCTCATTTCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGGACGGCC 2084
Db 1985 GCACACTGTCTACATGACAGAGGATTCAGGTACTCAGACCCACAGAGTGGACGGCC 2044
QY 2085 TTCTGGGTCTGTGACACAGAGCGAGTGGCTTGGGGTGGAGGAGTGAAGTTC 2144
Db 2045 TTCTGGGTCTGTGACACAGAGCGAGCGCGGCGGCTGGGGTGGAGGAGCGAGGCTC 2104
QY 2145 AAAAGTATCATACCCACCCCTTCTTCAATGATTTTCACTTTCAGCTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTTCACCTTTCAGCTATGACATCGCCTG 2164
QY 2205 CTGGAGCTGGAGAGTGTGGTGGAGTACAGACCGTGTGGCGCCCATCTGCTGCTGTGAT 2264
Db 2165 CTGGAGCTGGAGAGTGTGGTGGAGTACAGTCCATGCTGGCGGCGGCTTCTGCTGCGGAC 2224
QY 2265 GCTTACCCATCTTCTCTGTCGCAAGGCCATCTGGGTACAGGCTGGGGGCGACACAAA 2324
Db 2225 GCTTCCCATCTTCTCTGTCGCGCAAGGCCATCTGGGTACGGGCTGGGGACACCCAG 2284
QY 2325 GAGGAGGATACCGAGCGCTGTATCTTGCAGAAAGGTTGAGATCCGTGTCTATCAACAGAGC 2384
Db 2285 TATGGAGGCACTGGCGCTGTATCTTGCAGAAAGGTTGAGATCCGCGTCTATCAACAGAGC 2344
QY 2385 ACCTGTGAGAGCTCATGCGGAGAGATCAACCCAGAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAGCTCATGCGGAGAGATCAACCCAGAGATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGACCTTCCGACAGGCTGACTCTGGTGGCGGCTTGTCAAGCGCGAGAAA 2504
Db 2405 AGCGCGGCTGTGACTCTTCCAGGGGTGATTTCCGGGGGACCCCTGCTCCAGCGTGGAGCG 2464

```
QY 2505 GATGGCGAATGTTCCAGGCTGGTGTGAGAGCTGGGTGAAAGGTGGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATCTTCCAGGCGGTGTGAGTGGGAGAGCGGTGGCTCAGAGGAAC 2524
QY 2565 AAGCCAGGCGTGTACACAGAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGG 2624
Db 2525 AAGCCAGGCGTGTACACAGAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGGAACACTGGG 2584
QY 2625 GTATAGCAGCTGAGAGACAGCCGACCAACACACACCCACAGGAGTGCCTCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG--GCCACCCAAATGTGTACACCTGCGGGGCCACCCATGTCCTACC 2641
QY 2685 CTTGATACAGAGAGGAGACACTGAGCAATTTAGTGTGGCTTCCCTCCCTCCCTCCCTCC 2744
Db 2642 CCAGTGTGCACG-CCTGCGAGGTGGAGACTGACCCGCTGACTGCACCAAGCGGCC-CCAGA 2699
QY 2745 ACCGAGACTGTGAACCTGCATCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 15
US-10-112-221A-3
; Sequence 3, Application US/1011221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-112-221A-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGAGGAGCATCGGGCCCGCAAGGGGAGGGGGCCCGAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAAATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCAAGAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGAGTTCCTGCTGCGAACHATGCCAAGAAGTGGAGAAGGAGGAGGAGGAGGAGG 224
Db 125 GGGCGTGGAGTTCCTGCGAGTCAACACGTCAGAAAGGTGAAAAGCATGGCCCGGGGCGC 184
QY 225 TGGGTGGTCTGGTGGCAGTGTGTTTACGTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGGTCTGGCAGCGGTGTGATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGGTGTGCACTTCCATATACGAAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGCAATTTGCAATACCGGAGCGTGGCTGTCCAGAAAGTCTTCAATGGCTACATG 304
```

```
QY 345 AGGATCAAAATGATGATCTTTTGTGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGAAATTTTGTGATGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
QY 405 AGCTTGGCGCAGCAGGCTGAAGAGGCGCTGAAGCTGTGTGTACAAATGAAGTCCCTTCCTG 464
Db 365 AGCTTGGCGCAGCAAGGCTGAAGAGCAGCGCTGAAGCTGTGTGTACAGCGGAGTCCCATTC 424
QY 465 GGTCCCTTACCAACAAAGCTCGGCTGTAACTGCTTTCAGTGGGGCAGTGTCTATCCCTTAC 524
Db 425 GGCCTTACCAACAAAGGCTCGGCTGTGAAGCTTTCAGGAGGCGAGTGTCTATCCCTTAC 484
QY 525 TACTGGTCAGAGTTCAGCATCCCTCCCAACACTGGCAGAGAGGTTGATGCGGCCATGGCT 584
Db 485 TACTGGTCAGAGTTCAGCATCCCGCAGCACTGTGTGAGGAGGCGCGCTCATGTGCC 544
QY 585 GTGAGCGGAGTGTAAACATTCGCCACCCCGCAGCAGCGGCACTGAAATCTCTCGTGCTACA 644
Db 545 GAGGAGCGGCTAGTCACTGCTGCCCGCGCGCGCTCCCTGAAAGTCTTTGTGTGCTACC 604
QY 645 TCTGTGTGGCTTCCCTCAATTCACCCAGAAATGCTGCAGAGGACTCAGACAAACAGCTGC 704
Db 605 TCAGTGTGGCTTTCCTCCACCGACTCCAAACAGTACAGAGGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTGGCTGCTGATGCCCATGCTGTCAGCAGTGCACACGCTTCACTTACCCCTGGCTTCC 764
Db 665 AGCTTGGCTGCTGACGCGCGCGGTGTGAGAGTGTATGCTTCACTTACCCACGCGCTTCC 724
QY 765 AACAGTCTCCCTACCGGCGCATCGCGCTGCCAGTGGCTGCTGCGGGGAGCGCCAGCTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCGCGCTGCCAGTGGCGCGCTGCGGGGGAGCGCCAGCTCA 784
QY 825 GTGTGAGCTTCACTTCCGAAAGCTTTGATGTGCTCTCTGTGTGATGAGCATGGCAGTGCAC 884
Db 785 GTGTGAGCTTCACTTCCGAGCTTTCGACCTTGGCTCTCTGCGACAGCGCGGAGCGAGCGAC 844
QY 885 CTGTGACCGTGTATGATGAGCTGAGCCCGCATGAGAACCCGCTGCTGTGCTGGCTGTGT 944
Db 845 CTGTGAGCGGTGTACACACCTTGAAGCGGCTTGAAGCGGCGGCTGCTGTGAGTGTGT 904
QY 945 GGCACCTTCTCACCTCTCTCAACCTGACTTTTCTCTCTCTCCAGAACTCTTCTCTTCTGTC 1004
Db 905 GGCACCTTACCTCTCTCTCTCAACCTGACTTCTCTCTCTCTCCAGAACTCTCTGCTCATC 964
QY 1005 AGCTGATTAACCAATPACTGACCGCGCATCATCTGCTTTTGAAGCCACTTTCTTTCAGCTG 1064
Db 965 ACCTGATAACCAACACTGAGCGCGCATCCCGGCTTGAAGCGGCGGCTTCTTCTTCTTCT 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTGATGACACCCAGGAGCAATTTAGCAGCGCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGCGCGCTTACGTAAGCCCGAGGAGCAATTCACAGCGCC 1084
QY 1125 TACTATCCAGCGCACTTACCCCGCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGCGCACTTACCCCGCAACATCAACTGCACATGGAATATGAGGTGCCCAAC 1144
QY 1185 AACCGAAAGTGAAGGTGCTTCAAACTCTTCTATCTGTGTGAGCCCAACAGTACAGT 1244
Db 1145 AACCAGCATGTGAAGTGTGAGCTTCAAACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1204
QY 1245 GGCTCTGTCACCAAGGACTATGTGAGATCAACCGGAGAGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTTACGTGAGATCAATGGGAGAAATATTCGCGAGAGAGGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTTCACCTTCTTCTTCTTCTTCTTCT 1364
Db 1265 CAGTTCTGTGTACCAAGCAACAGCAACAGATCAAGTTCCGTTCTTCTTCTTCTTCTTCT 1324
QY 1365 TACAGGACACCGGTTCTTCTAGCTGAGTACCTCTCTTACGACTCCAAAGAGCCCGTCCCA 1424
Db 1325 TACACCGACACCGGCTTCTTGTAGTGAATACCTCTCTTACGACTCCAGTGAAGTCCCGG 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGGAGCGGTGCATCCGAAAGGAACTGCGCTGCGAGCGGTG 1484
```

Db 1385 GGGCAGTTACGTCGCGCACGGGCGGTGTATCCGAAGAGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGGGTACTTGGCGATGCAATGCCACCCACAGTTTC 1544
Db 1445 GCCGACTGCACCGACACAGCGATGAGTCTAACTGCACTTGGAGCGCGGCCACCAAGTTTC 1504
Qy 1545 ACGTGCAAAAACCAAGTTCTGCAAGCCCTTCTTGGGTCTGTGACAGTGTCAAGCACTGT 1604
Db 1505 ACGTGCAAGAACAAAGTTCTGCAAGCCCTTCTTGGGTCTGTGACAGTGTGAAACGACTGC 1564
Qy 1605 GGGACGGAAGTACGAGGAGGGCTGCACTGTCTCTGCTGGAGTTTCAAGTGTCCAAAT 1664
Db 1565 GGACACAACAGCGACGAGCGAGGGGTGCACTGTCTCGGCCCAAGACTTCAGGTGTCCAAAT 1624
Qy 1665 GGGAAAGTGTCTCCCTCTCAGAGCGAGAAAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTCTCGMAAAGCCAGCAGTGTCAATGGGAAGGACGACTGTGGGAGCGGTTC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTCTCTGCCCCCAAGGTGAACGTCTCACTTGTACCAAAACACCACTACCGCTGC 1744
Qy 1785 CAAATGGCTCTCTCTGACGAAAGGCAACCTGAGTGTGATGGGAAGACGAGCTGTAGC 1844
Db 1745 CTCAATGGCTCTCTCTGAGCAAGGGCAACCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAGGACTGTGCACTGTGGGTGCGGTCAATTCACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGCAAGAAATGCGGACGAGGGGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db 1865 GTTGTGGGGCACGGATGCGGATGAGGGCAGTGGGCCCTGGCAGGTAAAGCTGCATGCT 1924
Qy 1965 CTGGGCGAGGCGCACTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCGAGGCGCACATCTGGGGTGTCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTTCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAAGCAGTGGACGGCC 2044
Qy 2085 TTCTGGGTCTGTGACCGACGAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCTGGGGTTGCAGCACGACGAGCGAGCGAGCGCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAAGTATCATCACCCACCTTCTTCAATGATTTCACTTCACTTCACTGATGATGATGATGAT 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTCACTTCACTGATGATGATGATGAT 2164
Qy 2205 CTGGAGCTGGAGAAAGTGGTGGAGTACAGCACCGTCTGTGGGCCCACTTGGCTGCTTGTAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGCTCCATGTTGGTGGGCCCACTTGGCTGCGGAC 2224
Qy 2265 GCTACCATGTCTTCTCTGTGGCAAGGCCATCTGGGTGACAGCTGGGGGACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCTCTGTGGCAAGGCCATCTGGGTGACAGCTGGGGGACACACCCAG 2284
Qy 2325 GAGGAGGTACCGGAGCGCTGATCTCTGAGAAAGGTGAGATCGGTGTCATCAACACGACCC 2384
Db 2285 TATGAGGCACTGGCGGTGATCTCTGAAAAGGTGAGATCCCGGTATCAACACGACCC 2344
Qy 2385 ACCTGTGAGGACCTATGCGCGACGAGATCACCCACGAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGCGAGAACCTCTCTGCGACGAGATCACCGCGCATGATGTGGTGGGTTCCTC 2404
Qy 2445 AGTGGGGTGTGAGTCTCTGCGAGGGTGACTCTGTGGTGGCCCTTGTCAAGCGCGGAAAA 2504
Db 2405 AGCGGCGCGTGGGACTCTCTGCGAGGGTGAATTCGGGGGGAACCCCTGTCCAGCGTGGAGCG 2464
Qy 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGGTGAAGGCTCGCTCAGAGAAC 2564

Db 2465 GATGGGCGGATCTTCCAGGCGGTGTGTGAGCTGGGGAGACGGCTCGCTCAGAGAAC 2524
Qy 2565 AAGCCAGGCTGTACACAAGGCTCCCTGTAGTTCCGGACTGGATCAAGAGACACACTGGG 2624
Db 2525 AAGCCAGGCTGTACACAAGGCTCCCTGTAGTTCCGGACTGGATCAAGAGAACACTGGG 2584
Qy 2625 GTATAGCAGCATGGACAGACAGCGGACACAAACACCCACAGGGATCCCGACATGCACA 2684
Db 2585 GTATAGGGGCGGG---GCCACCAATGTGTACCTGTGGGGCCACCCATCGTCCACC 2641
Qy 2685 CCTGGATACAGGAGAGAAACACTGACGACATTTATGTGTGGCTTCCCCCCCCCAACACA 2744
Db 2642 CCAGTGTGCAGC-CCTGCAAGGCTGGAGACTGGACCGCTGACTGCACCGGCGCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAACCTGCATCCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

Search completed: February 21, 2004, 10:23:01
Job time : 1070 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:06:37 ; Search time 6211 Seconds
(without alignments)
12154.195 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catgtagacggtcccg.....ttaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gssprt.*
21: em_gssfun.*
22: em_gssmam.*
23: em_gssmus.*
24: em_gsspro.*
25: em_gssrod.*
26: em_gssphg.*
27: em_gssvrl.*
28: gb_gsel.*
29: gb_gss2.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3102.8	99.9	3248	11	BC008514 Mus muscu
2	3087.6	99.4	4017	11	AK052738 Mus muscu
3	956.4	30.8	2689	11	AK085410 Mus muscu
4	843.8	27.2	944	13	BQ885551 AGENCOURT

5	833.6	26.8	884	12	BI688367
6	788.8	25.4	803	12	BI690615
7	778.2	25.1	990	13	BQ951529
8	760.4	24.5	762	12	BQ919617
9	759.4	24.4	874	14	CB208872
10	747.2	24.1	798	12	BI664081
11	739.2	23.8	746	14	BY756896
12	739.2	23.8	846	12	BI416218
13	735.8	23.7	1116	10	BF301347
14	732	23.6	786	12	BI646380
15	721	23.2	928	12	BQ915045
16	720.8	23.2	777	12	BI150180
17	716	23.1	928	12	BI694441
18	710.8	22.9	838	12	BI656113
19	707.4	22.8	767	12	BI695744
20	706.8	22.8	840	12	BI453178
21	706.2	22.7	939	14	CA976536
22	692.6	22.3	725	12	BI248406
23	689.6	22.2	960	12	BI414245
24	687.2	22.1	853	10	BF540669
25	684.4	22.0	697	12	BI853355
26	681.8	22.0	708	12	BI078327
27	681.2	21.9	838	13	BU152438
28	678.2	21.8	1035	10	BF119461
29	671	21.6	671	12	BG861619
30	665.4	21.4	836	13	BU517436
31	658.6	21.2	728	12	BI112917
32	650.4	20.9	927	10	BF578780
33	647.4	20.8	649	12	BI697395
34	643.8	20.7	698	14	BY756283
35	633.6	20.4	886	12	BI657843
36	630.4	20.3	924	12	BI454477
37	623.8	20.1	748	10	BF584458
38	623.4	20.1	854	12	BI416184
39	621.4	20.0	765	10	BF118535
40	617.2	19.9	630	10	BF320582
41	613	19.7	613	14	CA551053
42	612.4	19.7	868	13	BU179919
43	610.8	19.7	698	12	BG870626
44	610.4	19.7	802	12	BI149332
45	608.8	19.6	790	12	BG968045

ALIGNMENTS

RESULT 1
BC008514
LOCUS
DEFINITION
Mus musculus, suppression of tumorigenicity 14 (colon carcinoma), clone IMAGE:3489914, mRNA.
ACCESSION
BC008514
VERSION
BC008514.1
KEYWORDS
HTC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3248)
Strausberg, R.
Direct Submission
Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsb-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 7 Row: h Column: 16
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6677898
 This clone has the following problem: no 5' EST match.

FEATURES

Location/Qualifiers
 1..3248
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CS7BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3489914"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue"
 /clone_lib="NCI CGAP Mam5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 734 a 918 c 928 g 668 t

BASE COUNT
 ORIGIN

Query Match 99.9%; Score 3102.8; DB 11; Length 3248;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	CATGTTAGACGGCTGCCGGAGGACACACGGTCTGAGACGGGATCGGACCCGAA	60
Db	111	CATGTTAGACGGCTGCCGGAGGACACACGGTCTGAGACGGGATCGGACCCGAA	170
Qy	61	CCATGGGTAGCAATCGGGGCGCAAGCGCGAGGGGCTCTCAGGACTTCGGCGGGAC	120
Db	171	CCATGGGTAGCAATCGGGGCGCAAGCGCGAGGGGCTCTCAGGACTTCGGCGGGAC	230
Qy	121	TCAAGTCAACTCCCGGTAGAGACATGATGGCTTGAGAGGGTGTGAGTTCCTGC	180
Db	231	TCAAGTCAACTCCCGGTAGAGACATGATGGCTTGAGAGGGTGTGAGTTCCTGC	290
Qy	181	CTGGCAACAATGCCAAGAAATGGAGAGCGAGGCCCGAGCGCTGGTGTGCTGGTG	240
Db	291	CTGGCAACAATGCCAAGAAATGGAGAGCGAGGCCCGAGCGCTGGTGTGCTGGTG	350
Qy	241	CAGTGTGTTGAGCTTCTCTGCTCTCCCTCATGGTGGCTTCTGTTGGCACTTCC	300
Db	351	CAGTGTGTTGAGCTTCTCTGCTCTCCCTCATGGTGGCTTCTGTTGGCACTTCC	410
Qy	301	ATTATCGGAATGTCGGGTTCAAAAGTCTTCAATGGCCATCTGAGATCAAAATGGA	360
Db	411	ATTATCGGAATGTCGGGTTCAAAAGTCTTCAATGGCCATCTGAGATCAAAATGGA	470
Qy	361	TCTTTCTGGATGGTATGAGAACTCCACCTCCACAGAGTTTATCAGGCTGGCCAGCAGG	420
Db	471	TCTTTCTGGATGGTATGAGAACTCCACCTCCACAGAGTTTATCAGGCTGGCCAGCAGG	530
Qy	421	TGAGAGGGGCTGAAGCTGTGTACAAATGAAGTCCCTGCTGGTCCCTTACCACAAGA	480
Db	531	TGAGAGGGGCTGAAGCTGTGTACAAATGAAGTCCCTGCTGGTCCCTTACCACAAGA	590
Qy	481	AGTGGGTGAATGCTTCAAGTGGGCGAGTGTATCGGCTACTACTGTGTCAGATTCA	540
Db	591	AGTGGGTGAATGCTTCAAGTGGGCGAGTGTATCGGCTACTACTGTGTCAGATTCA	650
Qy	541	GCATCCCGCCACCTGGCAGAGAGTGTATCGGCTACTACTGTGTCAGATTCA	600
Db	651	GCATCCCGCCACCTGGCAGAGAGTGTATCGGCTACTACTGTGTCAGATTCA	710
Qy	601	CATTGCCACCCGAGCAGGGCACTGAATCTTCTGCTGTAAACATCTGTGGTGGCTTCC	660
Db	711	CATTGCCACCCGAGCAGGGCACTGAATCTTCTGCTGTAAACATCTGTGGTGGCTTCC	770

Qy	661	CCATTGACCCAGCAATCTGCAGAGGACTCAGGACAAACAGCTGCAGTTTGCCTTGCATG	720
Db	771	CCATTGACCCAGCAATCTGCAGAGGACTCAGGACAAACAGCTGCAGTTTGCCTTGCATG	830
Qy	721	CCCATGFGTCAGCTGACACGCTTCTACTACCCCTGGCTTCCCAACAGTCCCTACCCGG	780
Db	831	CCCATGFGTCAGCTGACACGCTTCTACTACCCCTGGCTTCCCAACAGTCCCTACCCGG	890
Qy	781	CGCATGCCGCTGCGAGTGGTCTGCGGGGGAGCGGACCTCTGTGTGAGCTCAGCT	840
Db	891	CGCATGCCGCTGCGAGTGGTCTGCGGGGGAGCGGACCTCTGTGTGAGCTCAGCT	950
Qy	841	TCCGAAGCTTTGATGTGCTCCCTGTGATGAGCATGGCAGTGAACCTGTGTACCTGATG	900
Db	951	TCCGAAGCTTTGATGTGCTCCCTGTGATGAGCATGGCAGTGAACCTGTGTACCTGATG	1010
Qy	901	ATAGCCTGAGCCCCCATGGAACCCCAACGCTGTGTGTGCGGCTGTGTGGCACTTCTCACCT	960
Db	1011	ATAGCCTGAGCCCCCATGGAACCCCAACGCTGTGTGTGCGGCTGTGTGGCACTTCTCACCT	1070
Qy	961	CCTACAACTGACCTTCTCTCTCCCGAGACGCTTCTTCTTCCAGCTGCCAAGATGAGCAGT	1020
Db	1071	CCTACAACTGACCTTCTCTCTCCCGAGACGCTTCTTCTTCCAGCTGCCAAGATGAGCAGT	1130
Qy	1021	CTGACCCGCGACATCTCTGGCTTTGAGGCCACTTCTTCCAGCTGCCAAGATGAGCAGT	1080
Db	1131	CTGACCCGCGACATCTCTGGCTTTGAGGCCACTTCTTCCAGCTGCCAAGATGAGCAGT	1190
Qy	1081	GTGGCGCTTTTGTAGTGACACCAACGAGCATTTAGAGGCCCTTCTATCCAGGCCACT	1140
Db	1191	GTGGCGCTTTTGTAGTGACACCAACGAGCATTTAGAGGCCCTTCTATCCAGGCCACT	1250
Qy	1141	ACCGCCCAACATCAACTGCACATGGAATCAAGTGCACCAACCGGAACTGGAAG	1200
Db	1251	ACCGCCCAACATCAACTGCACATGGAATCAAGTGCACCAACCGGAACTGGAAG	1310
Qy	1201	TGGCTTCAAACTTCTTCTATCTGTGGACCCCAACGCTTACAGTGGGCTCTCTGACCAAGG	1260
Db	1311	TGGCTTCAAACTTCTTCTATCTGTGGACCCCAACGCTTACAGTGGGCTCTCTGACCAAGG	1370
Qy	1261	ACTATGTGGAGATCAACGGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTGTGAGCA	1320
Db	1371	ACTATGTGGAGATCAACGGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTGTGAGCA	1430
Qy	1321	GCAACAGCAGCAAGATTACAGTCCACTTCCATTTCTGATCACTCGTACACGACACCGGCT	1380
Db	1431	GCAACAGCAGCAAGATTACAGTCCACTTCCATTTCTGATCACTCGTACACGACACCGGCT	1490
Qy	1381	TCCTAGCTGAGTACCTCTCTTCTAGCACTCCAAACGCGGTGCCAGGAGTTTCAATGTGA	1440
Db	1491	TCCTAGCTGAGTACCTCTCTTCTAGCACTCCAAACGCGGTGCCAGGAGTTTCAATGTGA	1550
Qy	1441	AGACTGGACGCTCATCCGAAAGAACTGCGCTGCGACGCTGGGAGACTGCGCGATT	1500
Db	1551	AGACTGGACGCTCATCCGAAAGAACTGCGCTGCGACGCTGGGAGACTGCGCGATT	1610
Qy	1501	ATAGTATGAGCGCTTACTGCGCATGCAATGCCACCAACAGTTCAGTGCAGAAACAGT	1560
Db	1611	ATAGTATGAGCGCTTACTGCGCATGCAATGCCACCAACAGTTCAGTGCAGAAACAGT	1670
Qy	1561	TCGTGAAGCCCTTCTTCTGTGGTCTGTGACAGTGTCAACGCTGTGGGAGCGGAAGTGA	1620
Db	1671	TCGTGAAGCCCTTCTTCTGTGGTCTGTGACAGTGTCAACGCTGTGGGAGCGGAAGTGA	1730
Qy	1621	AGGAGGCTGCAGCTCTCTCTGCTGGGAGTTTCAAGTGTCCATGCGAGTGTCTCCCTC	1680
Db	1731	AGGAGGCTGCAGCTCTCTCTGCTGGGAGTTTCAAGTGTCCATGCGAGTGTCTCCCTC	1790
Qy	1681	AGAGCCAGAAAGTGAATGGGAAGGACAACTGTGGAGATGGGCTGTGACGAGGCTTCAATGTG	1740
Db	1791	AGAGCCAGAAAGTGAATGGGAAGGACAACTGTGGAGATGGGCTGTGACGAGGCTTCAATGTG	1850

1981	TGTTGGGGCTCGCTCATCTCTCTGACCTGGCTGGTCTCTGACGTCTATGCTTTTCAGG	2041
2092	TGTTGGGGCTCGCTCATCTCTCTGACCTGGCTGGTCTCTGACGTCTATGCTTTTCAGG	2151
2041	ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCTTCTCGGTCTGCTGG	2100
2152	ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCTTCTCGGTCTGCTGG	2211
2101	ACGAGAGACGGCAGTGGCTCTTGGGGTGCAGGAGCTGAAGCTCAAAACGTATCATCACCC	2160
2212	ACCAGACAAGCGCAGTGGCTCTTGGGGTGCAGGAGCTGAAGCTCAAAACGTATCATCACCC	2271
2161	ACCCTTCTTTCAATGATTTCACTTTCAGCTATGACATCGCTTGTGCTGGAGCTGGAGAGT	2220
2272	ACCCTTCTTTCAATGATTTCACTTTCAGCTATGACATCGCTTGTGCTGGAGCTGGAGAGT	2331
2221	CGTGGAGTACAGACCGTCTGTGGGCCCATCTGCCCTGCTGATGCTACCAATGTCTTCC	2280
2332	CGTGGAGTACAGACCGTCTGTGGGCCCATCTGCCCTGCTGATGCTACCAATGTCTTCC	239
2281	CTGCTGCACAGGCCATCTGGGTTCACAGGTGGGGGCACACAAAGAGGAGGTACCGGAG	2340
2392	CTGCTGCACAGGCCATCTGGGTTCACAGGTGGGGGCACACAAAGAGGAGGTACCGGAG	2451
2341	CGTGTATCTTCGACAAAGGGTGAGATCCGTGTCTATCAACACAGACCACTGTGAGGACTCA	2400
2452	CGTGTATCTTCGACAAAGGGTGAGATCCGTGTCTATCAACACAGACCACTGTGAGGACTCA	2511
2401	TGCGGACAGCAGATCACCCACAGAAATGATGTGTGGGTTCCTCAGTGGGGTGTGGACT	2460
2512	TGCGGACAGCAGATCACCCACAGAAATGATGTGTGGGTTCCTCAGTGGGGTGTGGACT	2571
2461	CTTGCCAGGGTGACTCTGTGGGCCCTTGTCAAGCGCAGAGAAAGATGGCGAAATGTTCC	2520
2572	CTTGCCAGGGTGACTCTGTGGGCCCTTGTCAAGCGCAGAGAAAGATGGCGAAATGTTCC	2631
2521	AGGCTGGTGTGTGAGCTGGGGTGAAGGTGCGCTCAGAGGAAACAGCCAGCGTGTACA	2580
2632	AGGCTGGTGTGTGAGCTGGGGTGAAGGTGCGCTCAGAGGAAACAGCCAGCGTGTACA	2691
2581	CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCAGACTGGGGTATAGCAGCTGAC	2640
2692	CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCAGACTGGGGTATAGCAGCTGAC	2751
2641	AGACGCCGACCAACAAACCCACAGGGATGCCGACATGACACACTGGATACAGAGAG	2700
2752	AGACGCCGACCAACAAACCCACAGGGATGCCGACATGACACACTGGATACAGAGAG	2811
2701	GAACACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACACCCAGACTGTGAAT	2760
2812	GAACACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACACCCAGACTGTGAAT	2871
2761	GCATCTTTAGGACTCAGAGTCTTCCAAAGTGGGACCCCTCAAGAGTTGGAGAGAGAAC	2820
2872	GCATCTTTAGGACTCAGAGTCTTCCAAAGTGGGACCCCTCAAGAGTTGGAGAGAGAAC	2931
2821	TTGCGTGTAGCGGCCAGCTGGGGGCAAGGTTTGATGGCAGCTTCCCGCTCTAGCC	2880
2932	TTGCGTGTAGCGGCCAGCTGGGGGCAAGGTTTGATGGCAGCTTCCCGCTCTAGCC	2991
2881	CTGAGCTGGGTCAAGATGATGTGTCCGGAGAGCTGCTTCCAACTGTCAATTGAGCTCCC	2940
2992	CTGAGCTGGGTCAAGATGATGTGTCCGGAGAGCTGCTTCCAACTGTCAATTGAGCTCCC	3051
2941	GGAGCCCTTATGGGAGGAGGGCTCAGGGTCACTCTTTTCAGGAGCGCCAGCCCTAGGA	3000
3052	GGAGCCCTTATGGGAGGAGGGCTCAGGGTCACTCTTTTCAGGAGCGCCAGCCCTAGGA	3111
3001	ACCCAGAAAGAGTGTACCTAAGGCTGAAATGTTTTGCTGTGTGCAGGGGTGGGTAT	3060
3112	ACCCAGAAAGAGTGTACCTAAGGCTGAAATGTTTTGCTGTGTGCAGGGGTGGGTAT	3171
3061	TTGAGAGTAAACATTTTATTTCTTTTAAAAA	3094

Db 3172 TTGAGTAAACATTTATTCTTTTAGCTAA 3205

RESULT 3

LOCUS

AK085410

DEFINITION

Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D630022P10 product:suppression of tumorigenicity 14 (colon carcinoma), full insert sequence.

ACCESSION

AK085410

VERSION

AK085410.1 GI:26102691

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus

ORGANISM

Mus musculus

REFERENCE

1

AUTHORS

Carninci, P. and Hayashizaki, Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Mech. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

2

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374

PUBMED

11042159

REFERENCE

3

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE

20530913

PUBMED

11076861

REFERENCE

4

AUTHORS

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kueh, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Anono, H., Baldarelli, R., Baxsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.

TITLE

Functional annotation of a full-length mouse cDNA collection

JOURNAL

Nature 409 (6821), 685-690 (2001)

MEDLINE

21085660

PUBMED

11217851

REFERENCE

5

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & II Team.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-APR-2002)

Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

FEATURES

Location/Qualifiers

1..2689

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:D630022P10"

/db_xref="taxon:10090"

/clone="D630022P10"

/tissue type="kidney"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev stage="0 day neonate"

1..2689

/notes="suppression of tumorigenicity 14 (colon carcinoma) (MGD|MG1:133881, GB|NM_011176, evidence: BLASTN, 100%, match=720)"

misc_feature

590 a 747 c 766 g 586 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1415; Conservative

0; Mismatches 1; Indels 428; Gaps 3;

QY 1 CATGTAGACGGCTGCGCGAGGAGCACCGCTCTGAGACCGCGCATCGACCGCAAA 60

DB 110 CATGTAGACGGCTGCGCGAGGAGCACCGCTCTGAGACCGCGCATCGACCGCAAA 169

QY 61 CCATGGGTAGCAATCGGGCCCGCAAGCGCGAGGCGGCTCTCAGGACTTCGGCGCGGAC 120

DB 170 CCATGGGTAGCAATCGGGCCCGCAAGCGCGAGGCGGCTCTCAGGACTTCGGCGCGGAC 229

QY 121 TCNAGTACAACTCCCGGTAGAGACATGAATGGCTTTGAGGAGGGTGGAGTTCCTGC 180

DB 230 TCAAGTACAACTCCCGGTAGAGACATGAATGGCTTTGAGGAGGGTGGAGTTCCTGC 289

QY 181 CTGCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGCGCCAGGCGCTGGGTGGTGGTGG 240

DB 290 CTGCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGCGCCAGGCGCTGGGTGGTGGTGG 349

QY 241 CAGTGTGTTAGCTTCCTTTCTCTCTCCATGGCTGGCTTGTGTGTGGCACTTCC 300

DB 350 CAGTGTGTTAGCTTCCTTTCTCTCTCCATGGCTGGCTTGTGTGTGGCACTTCC 409

QY 301 ATTATCGGATGTGGGGTTCAAAAAGTCTCAATGGCCATCTGAGGATCACAATGAGA 360

```

410 ATTATCGGAATGTGGGGTCAAAAGTCTTCAATGGCCATCTGAGATCAAAATGAGA 469
361 TCTTTCTGGATGCGPATGAGAACTCCACTCCACAGAGTTTATCAG-CTGGCCAGCCAG 419
470 TCTTTCTGGATGCGPATGAGAACTCCACTCCACAGAGTTTATCAGCCCTGGCCAGCCAG 529
420 GTGAAGAGGGGCTGAAGTGTGTACAAATGAGTCCCTGCTGGGTCCTTACACAAG 479
530 GTGAAGAGGGGCTGAAGTGTGTACAAATGAGTCCCTGCTGGGTCCTTACACAAG 589
480 AAGTGGCTGTAACTGCTTCAGTGGAGGCGAGTGTATCGCTACTACTGTGTGAGTTC 539
590 AAGTGGCTGTAACTGCTTCAGTGGAGGCGAGTGTATCGCTACTACTGTGTGAGTTC 649
540 AGATTCCTCCACACCTCGCAGAGAGGTGTATCGGCCATGGCTGTGGAGCGAGTTGTA 599
650 AGCATCCCTCCACACCTCGCAGAGAGGTGTATCGGCCATGGCTGTGGAGCGAGTTGTA 709
600 ACATTGCCACCCGAGCAGCGGCACTGAATCTCTGCTGTGCTAACTGTGTGGTCCCTTC 659
710 ACATTGCCACCCGAGCAGCGGCACTGAATCTCTGCTGTGCTAACTGTGTGGTCCCTTC 769
660 ----- 659
770 CGTGAGTGGGGGCAACACGAGTGGGAATGTGTGGGCCAGCCTGTAAAGGGGCGGGCT 829
660 ----- 659
830 GGGAGAGCCACTCACAGGCCAGGTGTCTCGGAGGACTCTGCTGTGAGTGGCACT 889
660 ----- 659
890 CTTTGTGGGTAAATGGAGGCAAGAGTAAGAAATGATGTGTGGTAAATGTTTGCT 949
660 ----- 696
950 TCATGTTCTCCACCCTATGACCCAGAGTGTGAGAGGACTCAGGACAGTAATATCAT 1009
697 ----- 696
1010 GGTGACCTCTGGTGGAGAGAAAGTGTGTGCTTCCCTTAAGCCATGGGTCCCTAAAGT 1069
697 ----- 696
1070 CAGATGTAAGTGAAGCAAGCATCTCTCTTTTCCCTTCCGGTCCAAAGCTCTTTTC 1129
697 ----- 696
1130 TTGTGTCTACCTGAGTTCGAGTGTGATCTGAGACGTATTGGGGGTAGAGCTAGCGGGGG 1189
697 ----- 712
1190 GGGGGGCTGCGAGGTTTCAAGCTTCTCTTCCATCTCTCCACAGACAGTGTGAGTTTC 1249
713 CCTCATGCCCATGCTGCGAGTGTGACAGTGTGCTTCACTACCTGGCTTCCCAACAGTCC 772
1250 CCTCATGCCCATGCTGCGAGTGTGACAGTGTGCTTCACTACCTGGCTTCCCAACAGTCC 1309
773 CTACCCGGCGCATGCCGCTGCCAGTGGGTCTGCGGGGGAGCCGACTGTGCTGAG 832
1310 CTACCCGGCGCATGCCGCTGCCAGTGGGTCTGCGGGGGAGCCGACTGTGCTGAG 1369
833 CCTCACCTTCGGAAGCTTTGATGTGCTTCCCTGTGATGAGCATGGCAGTGTGTCAC 892
1370 CCTCACCTTCGGAAGCTTTGATGTGCTTCCCTGTGATGAGCATGGCAGTGTGTCAC 1429
893 CGTGTATGATAGCTGTAGCCCTTGAATGCTTCCCTGTGATGAGCATGGCAGTGTGTCACCTT 952
1430 CGTGTATGATAGCTGTAGCCCTTGAATGCTTCCCTGTGATGAGCATGGCAGTGTGTCACCTT 1489
953 CTCACCTCTCTACACCTGACTTCTCTCTCTCCAGAACTGCTTCTCTGTCAGCTGAT 1012

```

```

Db 1490 CTCACCTCTCTACAACTGACTTTCCTCTCTCTCCAGAACTCTCTCTGTCACGCTGAT 1549
QY 1013 AACCAATATCTACCGCGGACATCTGGCTTTGAGCCACTTTCTTCCAGCTGCCCAAGAT 1072
Db 1550 AACCAATATCTACCGCGGACATCTGGCTTTGAGCCACTTTCTTCCAGCTGCCCAAGAT 1609
QY 1073 GAGCAGCTGTGGCGCTTTTGTGAGTGACACCCAAAGGACATTTAGCAGCCCTCTATATCC 1132
Db 1610 GAGCAGCTGTGGCGCTTTTGTGAGTGACACCCAAAGGACATTTAGCAGCCCTCTATATCC 1669
QY 1133 AGGCATCTACCGCGGACATCTGGCTTTGAGTGACACCCAAAGGACATTTAGCAGCCCTCTATATCC 1192
Db 1670 AGGCATCTACCGCGGACATCTGGCTTTGAGTGACACCCAAAGGACATTTAGCAGCCCTCTATATCC 1729
QY 1193 CGTGAAGTGGCTTTCAAACTCTTCTATCTGCTGAGCCCAAGCTTACAGTGGGCTCTCTG 1252
Db 1730 CGTGAAGTGGCTTTCAAACTCTTCTATCTGCTGAGCCCAAGCTTACAGTGGGCTCTCTG 1789
QY 1253 CACCAAGGACTATGTGAGATCAACGGGAGAGTACTTCTGCTGAGAGTCTCCAGTTTGT 1312
Db 1790 CACCAAGGACTATGTGAGATCAACGGGAGAGTACTTCTGCTGAGAGTCTCCAGTTTGT 1849
QY 1313 GGTGAGCAGCAACAGCAGAGATTAAGTCCATCTTCAATCTGATCACTGTCACAGGA 1372
Db 1850 GGTGAGCAGCAACAGCAGAGATTAAGTCCATCTTCAATCTGATCACTGTCACAGGA 1909
QY 1373 CACCGGTTCTAGCTGAGTACCTCTCTAGCTTCCCAAGCACC 1416
Db 1910 CACCGGTTCTAGCTGAGTACCTCTCTAGCTTCCCAAGCACC 1953

RESULT 4
BQ885551
LOCUS
DEFINITION BQ885551 944 bp mRNA linear EST 16-AUG-2002
AGENCOURT 8748027 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335237
S, mRNA sequence.
ACCESSION BQ885551
VERSION BQ885551.1 GI:22277569
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 944)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13797 row: f column: 06
High quality sequence stop: 623.
FEATURES
Location/Qualifiers
1..944
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335237"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 130"
/notes="Organ: otcyats; Vector: pCMV-SPORT6.1.ccdh;
Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 184 a 288 c 272 g 200 t

```


Db 361 GCAATGCCACCCAGGTTACGTGCAAAACAGGTTCTGCAAGCCCTCTCTGGGTCT 420
QY 1585 GTGACAGTGTCAACGACTGTGGGGAGGAGTGACGAGAGGCTGCGAGCTCTCTGCTG 1644
Db 421 GTGACAGTGTCAACGACTGTGGGGAGGAGTGACGAGAGGCTGCGAGCTCTCTGCTG 480
QY 1645 GGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTCAGAGCCAGAAAGTGAATGGAAGG 1704
Db 481 GGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTCAGAGCCAGAAAGTGAATGGAAGG 540
QY 1705 ACAACTGTGGAGATGGTCTGACGAGGCTTCATGTGACAGCTGATGTCTCTTGA 1764
Db 541 ACAACTGTGGAGATGGTCTGACGAGGCTTCATGTGACAGCTGATGTCTCTTGA 600
QY 1765 CCAATATATACCTACCGCTGCCAAATGGCTCTGTCTGAG-CAAGGGCAACCTCAGTGT 1823
Db 601 CCAATATATACCTACCGCTGCCAAATGGCTCTGTCTGAGCCAGGGCAACCTCAGTGT 660
QY 1824 GATGGAGACGAGCTGTAGCCATGGCTCCGATGAGAAACCTGTGACTGTGGCTGCGA 1883
Db 661 GATGGAGACGAGCTGTAGCCATGGCTCCGATGAGAAACCTGTGACTGTGGCTGCGA 720
QY 1884 TCCCTTACCAACAGCTCGCG-TGGTGGTGGCAGGATGCGGACGAGGCGAGTGCC 1942
Db 721 TCCCTTACCAACAGCTCGCGTTGGTGGCAGGATGCGGACGAGGCGAGTGCC 780
QY 1943 CTGGCAGTGAGCTTCCAGCCCTGGGCGAGGCGACCTTGTGTGGGGCTCGC-TCATCT 2001
Db 781 CTGGCAGTGAGCTTCCAGCCCTGGGCGAGGCGACCTTGTGTGGGGCTCGCTCATCT 840
QY 2002 CTCCTGACTGGCTGTCTGTGAGCTCATTTGCTTTCAGGATGAC 2045
Db 841 CTCCTGACTGGCTGTCTGTGAGCTCATTTGCTTTCAGGATGAC 884

RESULT 6

BI690615 803 bp mRNA linear EST 18-SEP-2001
LOCUS 603314210F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5354067 5',
DEFINITION mRNA sequence.

ACCESSION BI690615

VERSION BI690615.1 GI:15653244

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 803)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM11900 row: d column: 04

High quality sequence stop: 755.

Location/Qualifiers

1. 803

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5354067"

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

FEATURES

Source

/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 170 a 259 c 200 g 174 t

Query Match

Best Local Similarity 25.4%; Score 788.8; DB 12; Length 803;

Matches 801; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 654 GCCTTCCCATTTGACCCAGAAATGCTGCAGAGGACTCAGGACCAACAGCTGCGATTTGCC 713
Db 1 GCCTTCCCATTTGACCCAGAAATGCTGCAGAGGACTCAGGACCAACAGCTGCGATTTGCC 60
QY 714 CTGATGCCCATTTGATGTCAGCAGTGCACGCTTCACTACCCCTGGCTTCCCAACAGTCCC 773
Db 61 CTGATGCCCATTTGATGTCAGCAGTGCACGCTTCACTACCCCTGGCTTCCCAACAGTCCC 120
QY 774 TACCGGGCGATGCGCGTGCAGTGGTCTCTGCGGGGAGCGCGACTCTGTGTGAGC 833
Db 121 TACCGGGCGATGCGCGTGCAGTGGTCTCTGCGGGGAGCGCGACTCTGTGTGAGC 180
QY 834 CTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGCCTGTCACC 893
Db 181 CTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGCCTGTCACC 240
QY 894 GTGATGATAGCTGAGCCCATGGAACCCACGCTGTGTGCGGCTGTGTGGACCTTC 953
Db 241 GTGATGATAGCTGAGCCCATGGAACCCACGCTGTGTGCGGCTGTGTGGACCTTC 300
QY 954 TCACCTTCTACAACTGACTTTCTCTCTCCAGAAAGTCTCTTGTCAAGCTGATA 1013
Db 301 TCACCTTCTACAACTGACTTTCTCTCTCCAGAAAGTCTCTTGTCAAGCTGATA 360
QY 1014 ACCAATGCTGACCGCGGACATCTCTGAGGCTTTGAGGCGCACTTCTTCCAGTCCCAAGATG 1073
Db 361 ACCAATGCTGACCGCGGACATCTCTGAGGCTTTGAGGCGCACTTCTTCCAGTCCCAAGATG 420
QY 1074 AGCAGCTGTGGCGCTTTTGGTGTGACACCCCAAGGGACATTTAGCAGCCCTTCTATCCA 1133
Db 421 AGCAGCTGTGGCGCTTTTGGTGTGACACCCCAAGGGACATTTAGCAGCCCTTCTATCCA 480
QY 1134 GGCCACTACCGCGGACATCACTGACATGATGATCAAGTGCAGGACCAACCGGAAC 1193
Db 481 GGCCACTACCGCGGACATCACTGACATGATGATCAAGTGCAGGACCAACCGGAAC 540
QY 1194 GTGAAGTGTGGCTTCAAACTCTTCTATCTGTGTGACCCCAAGTACCAGTGGGTCTCTGC 1253
Db 541 GTGAAGTGTGGCTTCAAACTCTTCTATCTGTGTGACCCCAAGTACCAGTGGGTCTCTGC 600
QY 1254 ACCAAGCTATGTGGAGATCAACGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTG 1313
Db 601 ACCAAGCTATGTGGAGATCAACGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTG 660
QY 1314 GTGAGCAGCAACGACGACAGATGATGATCCACTTCCATTCTGATCACTGTACACGGAC 1373
Db 661 GTGAGCAGCAACGACGACAGATGATGATCCACTTCCATTCTGATCACTGTACACGGAC 720
QY 1374 ACCGGTTCCTAGCTGAGTACCTCTCTCAGACTCCAAACGACCGTCCCAAGGATGTC 1433
Db 721 ACCGGTTCCTAGCTGAGTACCTCTCTCAGACTCCAAACGACCGTCCCAAGGATGTC 780
QY 1434 ATGTGCAAGTGGACGCTGCATC 1457
Db 781 ATGTGCAAGTGGACGCTGCATC 803

RESULT 7

BQ951529

LOCUS

BQ951529 990 bp mRNA linear EST 21-AUG-2002


```

DEFINITION AGENCOURT_8952916 NCI_CGAP_Co24 Mus musculus cDNA clone
ACCESSION IMAGE:6475074 5', mRNA sequence.
VERSION BQ951529
KEYWORDS BQ951529.1 GI:22367007
SOURCE EST.
ORGANISM Mus musculus (house mouse)
REFERENCE Mus musculus
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14013 row: d column: 19
High quality sequence stop: 635.
FEATURES
    Location/Qualifiers
        1..990
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="FVB/N"
            /db_xref="taxon:10090"
            /clone="IMAGE:6475074"
            /lab_host="DH10B (TI phase-resistant)"
            /clone_lib="NCI_CGAP_Co24"
            /note="Organ: Colon; Vector: pCMV-Sport6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 1.6 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 229 a 272 c 286 g 201 t 2 others
ORIGIN

Query Match 25.1%; Score 778.2; DB 13; Length 990;
Best Local Similarity 92.9%; Pred. No. 1.9e-142;
Matches 924; Conservative 0; Mismatches 60; Indels 11; Gaps 10;

QY 981 TCCTCCAGACGTCCTTGTGTCAGCTGATACCAATATGACCGGCGACATCTCTGGC 1040
Db 1 TCCTCCAGACGTCCTTGTGTCAGCTGATACCAATATGACCGGCGACATCTCTGGC 60
QY 1041 TTTGAGGCGCACTTCTTCCAGCTGCCAAGATGACGAGCTGTGGCGGCTTTTTCAGTGAC 1100
Db 61 TTTGAGGCGCACTTCTTCCAGCTGCCAAGATGACGAGCTGTGGCGGCTTTTTCAGTGAC 120
QY 1101 ACCCAAGGACATTTAGCAGCCCTTACTATCCAGGCCACTACCGGCCCAACATCAACTGC 1160
Db 121 ACCCAAGGACATTTAGCAGCCCTTACTATCCAGGCCACTACCGGCCCAACATCAACTGC 180
QY 1161 ACATGGATATCAGGTGCCCAACACCGGACCTGAGGTGCGCTTCAACTCTTCTAT 1220
Db 181 ACATGGATATCAGGTGCCCAACACCGGACCTGAGGTGCGCTTCAACTCTTCTAT 240
QY 1221 CTGTGTGACCCCAACGTTACAGTGGGCTCTCTGACCAAGGACTATGTGGAGATCAACGGG 1280
Db 241 CTGTGTGACCCCAACGTTACAGTGGGCTCTCTGACCAAGGACTATGTGGAGATCAACGGG 300
QY 1281 GAGAAGTACTCGGTGAGAGTCCAGTTTGTGTGAGCAGCAACAGCAGCAAGATTACA 1340
Db 301 GAGAAGTACTCGGTGAGAGTCCAGTTTGTGTGAGCAGCAACAGCAGCAAGATTACA 360
QY 1341 GTCCACTTCCATTTCTGATCACTCTGACGAGACCGGGTTCCTAGCTGAGTACCTCTCC 1400
Db 361 GTCCACTTCCATTTCTGATCACTCTGACGAGACCGGGTTCCTAGCTGAGTACCTCTCC 420
QY 1401 TACGACTCCAACGACCGGTGCCCGAGGATGTTTCATGTGCAAGACTGGACGGTGCATCCGA 1460

```

```

Db 421 TACGACTCCAACGACCGGTGCCCGAGGATGTTTCATGTGCAAGACTGGACGGTGCATCCGA 480
QY 1461 AAGGAATCGCTGCGACCGGTGGGCGAGACTGCCCGGATTATAGTAGAGCGTTACTGCG 1520
Db 481 AAGGAATCGCTGCGACCGGTGGGCGAGACTGCCCGGATTATAGTAGAGCGTTACTGCG 540
QY 1521 CGATGCAATGCCACCCACCACTGTTCCAGTGCAGAAACCAAGTTCGCAAGCCCT-CTTCTG 1579
Db 541 CGATGCAATGCCACCCACCACTGTTCCAGTGCAGAAACCAAGTTCGCAAGCCCTCTTCTG 600
QY 1580 GGTCTGTGACAGTGTCAACGACT-GTGGGACGGAAGTGCAGGAGGGTGCAGCTGCTC 1638
Db 601 GGTCTGTGACAGTGTCAACGACTGTTGGGACGCGAAGTGCAGGAGGGTGCAGCTGCTC 660
QY 1639 CTGCTGGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTCAGAGCAGAGT-GTAAT 1697
Db 661 CTGCTGGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTCAGAGCAGAGTGGTAAAT 720
QY 1698 GGAAGGACAACTGT--GGAGATGGGTCTGACGAGGCTTCATGTG-ACAGCGTGAATGTC 1754
Db 721 GGAAGGACAACTGTGGGAGATGGGTCTGACGAGGCTTCATGTGAAACGGTGAATGTC 780
QY 1755 GTCTTTGGACCAATATACCTACCGCTGCCAAATGGCTCTGTCTGAGCAAGGGCAAC 1814
Db 781 GTCTTTGGACCA-ATATATACCTACCGCTGCCAAATGG-CTCTGTCTGAGC-AGGGCACC 837
QY 1815 CCTGAGTGTGATGGAGACGGACTGTAGGATGGCTCCGATGAGTGAAGAACTGACTGT 1874
Db 838 CTGATTTGGATGGGAAGACGGACTGTACCGATGG-TCCGATGAAAAAAC-GGGAATGT 895
QY 1875 GGGTGTGATCTTTTACCAACAGGCTCGGTGTGTTGTGGCACGAATGGGACGAGGGC 1934
Db 896 GGGTGTGATCTTTTACCAACAGGCTCGGTGTGTTGTGGCACGAATGGGACGAGGGC 955
QY 1935 GAGTGGCTTGGGACGGTGCAGGTCAGCTTCCACGCCCTGGG 1969
Db 956 CAGGGGCGCGGTGAGGGGACCTCCCGGCGGTGG 990

RESULT 8
BQ919617
LOCUS 602823173F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4951962 5',
mRNA sequence.
ACCESSION BQ919617
VERSION BQ919617
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10909 row: a column: 19
High quality sequence stop: 727.
FEATURES
    Location/Qualifiers
        1..762
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="FVB/N"
            /db_xref="taxon:10090"

```

```
/clone="IMAGE:4951962"
/sex="female, virgin"
/tissue type="infiltrating ductal carcinoma"
/dev stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      164 a 243 c 190 g 165 t
ORIGIN
Query Match      24.5%; Score 760.4; DB 12; Length 762;
Best Local Similarity 99.9%; Pred. No. 6.4e-139;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 682 AGAGGACTCAGGACACAGCTGCAGTTTGGCCCTGCATGCCATGGTCAGCAGTGACAC 741
Db 1 AGAGGACTCAGGACACAGCTGCAGTTTGGCCCTGCATGCCATGGTCAGCAGTGACAC 60
QY 742 GCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGGCGCATCCCGCTGCCAGTGGG 801
Db 61 GCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGGCGCATCCCGCTGCCAGTGGG 120
QY 802 TCCTCGGGGGGACGCCGACTCTGTGTGAGCCTCACCTTCGGAAGCTTGATGTCGCTC 861
Db 121 TCCTCGGGGGGACGCCGACTCTGTGTGAGCCTCACCTTCGGAAGCTTGATGTCGCTC 180
QY 862 CTTGTGATGAGCATGGCAGTGCAGCTGTGCACCTGTATGATAGCCTGAGCCCATGGAA 921
Db 181 CTTGTGATGAGCATGGCAGTGCAGCTGTGCACCTGTATGATAGCCTGAGCCCATGGAA 240
QY 922 CCCAGCTGTGTGGCTGTGTGGCACTTCTCACCTCTTACCACTGACTTTCCTCT 981
Db 241 CCCAGCTGTGTGGCTGTGTGGCACTTCTCACCTCTTACCACTGACTTTCCTCT 300
QY 982 CTCCCGAAGCTCTTCTTGTGACGCTGATTAACCAATACTGACCGGCGACATCCTGCT 1041
Db 301 CTCCCGAAGCTCTTCTTGTGACGCTGATTAACCAATACTGACCGGCGACATCCTGCT 360
QY 1042 TTGAGGCCACTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACA 1101
Db 361 TTGAGGCCACTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACA 420
QY 1102 CCCAAGGACATTTAGCAGCCCTACTATCCAGGCCACTTACCGCCCAACATCAACTGCA 1161
Db 421 CCCAAGGACATTTAGCAGCCCTACTATCCAGGCCACTTACCGCCCAACATCAACTGCA 480
QY 1162 CATGGAATATCAAGTGGCCCAACACCGGAAGCTGGAAGTGGGCTTCAAACTCTTCTATC 1221
Db 481 CATGGAATATCAAGTGGCCCAACACCGGAAGCTGGAAGTGGGCTTCAAACTCTTCTATC 540
QY 1222 TGGTGACCCCAACCTTACAGTGGGCTCTCTGACCAAGACTATGTGGAGATCAACGGGG 1281
Db 541 TGGTGACCCCAACCTTACAGTGGGCTCTCTGACCAAGACTATGTGGAGATCAACGGGG 600
QY 1282 AGAAGTACTGCGGTGAGAGGTCCCAAGTTGTGGTGAGCAGCAACAGCAGCAAGATTACAG 1341
Db 601 AGAAGTACTGCGGTGAGAGGTCCCAAGTTGTGGTGAGCAGCAACAGCAGCAAGATTACAG 660
QY 1342 TCCACTTCGATCTGATCACTGTHACAGGACACCGGGTTCCTAGCTGAGTACCTCTCT 1401
Db 661 TCCACTTCGATCTGATCACTGTHACAGGACACCGGGTTCCTAGCTGAGTACCTCTCT 720
QY 1402 ACGACTTCCAAACGACCGTCCCGAGGATGTTTCATGTGCAAGA 1443
Db 721 ACGACTTCCAAACGACCGTCCCGAGGATGTTTCATGTGCAAGA 762
```

RESULT 9
CB208872
LOCUS

DEFINITION AGENCOURT 11350695 NIH MGC 164 Mus musculus cDNA clone
IMAGE:30241011 5', mRNA sequence.
ACCESSION CB208872
VERSION CB208872.2 GI:29133712
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 874)
NIH-MGC <http://mgi.nci.nih.gov/>
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT On Feb 4, 2003 this sequence version replaced gi:28250435.
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM0308 row: g column: 04
High quality sequence stop: 655.
Location/Qualifiers
1. 874
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clones="IMAGE:30241011"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 164"
/note="Vector: pCMV-SPORT6.1.ccd; Site 1: EcoRV; Site 2:
NotI; Non-normalized full-length enriched library from
pooled mouse embryonic limb, maxilla and mandible, day
10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp, Average insert size 1.8k bp. Priming
sequence: 5'-GACTAGTTCTAGATCGGAGCGCGCC(T) 3', Tissue
contributed by, David Rowe. Library constructed by ResGen,
Invitrogen Corp."

BASE COUNT 193 a 230 c 260 g 190 t 1 others

ORIGIN

Query Match 24.4%; Score 759.4; DB 14; Length 874;
Best Local Similarity 98.9%; Pred. No. 9.8e-139;
Matches 785; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
QY 1387 CTGAGTACCTCTCTACGACTCCAAACGACCCGTCGCCAGGATGTTTATGTGCAAGACTG 1446
Db 1 CTGAGTACCTCTCTACGACTCCAAACGACCCGTCGCCAGGATGTTTATGTGCAAGACTG 60
QY 1447 GACGCTGATCCGAAGAACTGGCTGCGAGCGCTGGCGAGACTGCCCGATTATAGTG 1506
Db 61 GACGCTGATCCGAAGAACTGGCTGCGAGCGCTGGCGAGACTGCCCGATTATAGTG 120
QY 1507 ATGAGCGTTTACTGCGGATGCAATGCCACCACTGTTTACGTGCAAAACCACTGTTCTGCA 1566
Db 121 ATGAGCGTTTACTGCGGATGCAATGCCACCACTGTTTACGTGCAAAACCACTGTTCTGCA 180
QY 1567 AGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGTGTGGGAGCGGAAGTGCAGGAGG 1626
Db 181 AGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGTGTGGGAGCGGAAGTGCAGGAGG 240
QY 1627 GCTCAGCTGTCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCC 1686
Db 241 GCTCAGCTGTCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCC 300
QY 1687 AGAAGTGTAAATGGGAAGGACAACTGTGGAGATGGTCTGACGAGGCTTCATGTGACAGCG 1746
Db 301 AGAAGTGTAAATGGGAAGGACAACTGTGGAGATGGTCTGACGAGGCTTCATGTGACAGCG 360

VERSION
KEYWORDS
SOURCE
ORGANISM

BY756896.1 GI:27190109

EST.
Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 746)

REFERENCE
AUTHORS

Okazaki, Y., Furuno, M., Saito, N., Osato, N., Hasegawa, Y., Suzuki, H., Yamanaka, I., Kiyosawa, H., Nikaido, I., Tomaru, Y., Tomaru, Y., Saito, N., Osato, N., Hasegawa, Y., Suzuki, H., Yamanaka, I., Kiyosawa, H., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., C.,

Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,

Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,

L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,

A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,

Gough, J., Grimmond, S., Gustinch, S., Hirokawa, N., Jackson, I. J.,

Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,

King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,

P. A., Maglott, D. R., Maitais, L., Marchionni, L., McKenzie, L., Miki,

H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perteau, G.,

Pesole, G., Petrovsky, N., Pallal, R., Pontius, J. U., Qi, D.,

Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,

B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C. A., Setou,

M., Shimada, K., Sulcane, R., Takenaka, Y., Taylor, M. S., Teasdale,

R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,

Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa,

M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A.,

Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,

M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,

Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,

Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,

K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,

E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Sueno-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,

S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,

Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,

H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,

Namazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N., Sano, H.,

Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,

Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

FEATURES
source

Location/Qualifiers
1..746

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="I1C0035F06"

/dev_stage="blastocyst"

/clone_lib="RIKEN full-length enriched, blastocyst"

172 a 193 c 224 g 155 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 741; Conservative

0; Mismatches 5; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

RESULT 12

BI416218

LOCUS

DEFINITION

602989878F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5145762 5',

linear EST 14-AUG-2001

BI416218

602989878F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5145762 5',

linear EST 14-AUG-2001

```

mRNA sequence.
ACCESSION B1416218
VERSION B1416218.1 GI:15177141
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1359 Row: 1 column: 19
High quality sequence start: 20
High quality sequence stop: 793.
Location/Qualifiers
1. .846
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5145762"
/tissue type="pooled lung tumors"
/lab host="DH10B (phage-resistant)"
/clone lib="NCI CGAP L33"
/note="Organ: Lung; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; Site: NotI; Site 2: EcoRI; lsc
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGAGCGGCGCTCTCTTTTCTTTTCTTTT
3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 178 a 224 c 262 g 182 t
ORIGIN
Query Match 23.8%; Score 739.2; DB 12; Length 846;
Best Local Similarity 97.1%; Pred. No. 9e-135;
Matches 816; Conservative 0; Mismatches 18; Indels 6; Gaps 6;
QY 1687 AGAAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 1746
DB 1 AGATCCGACAGATAGAGACACTGTGAGATGGTCTGACAGGCTTCATGTGACAGCG 60
QY 1747 TGAATGTCGTCTTTGACCAATAATACCTACCGCTGCCAAATGGCTCTGTCTGAGCA 1806
DB 61 TGAATGTCGTCTTTGACCAATAATACCTACCGCTGCCAAATGGCTCTGTCTGAGCA 120
QY 1807 AGGGCAACCTGAGTGTGATGGAGACGAGCTGTAGCGATGGCTCCGATGAGAAACT 1866
DB 121 AGGGCAACCTGAGTGTGATGGAGACGAGCTGTAGCGATGGCTCCGATGAGAAACT 180
QY 1867 GTGACTGTGGGCTGGATCTTTACAAACAGGCTCGGTGGTGGTGGGACGATGGG 1926
DB 181 GTGACTGTGGGCTGGATCTTTACAAACAGGCTCGGTGGTGGTGGGACGATGGG 240
QY 1927 ACGAGGCGAGTGGCCCTGGCAGGTGAGCTTCACGCCCTGGCCAGGGCCACTTGTGTG 1986
DB 241 ACGAGGCGAGTGGCCCTGGCAGGTGAGCTTCACGCCCTGGCCCA-GGCCACTTGTGTG 299
QY 1987 GGGCCCTCGCTCATCTCTCTCTGACTGGTGGTCTCTGACGACTCATGTCTTCAGGATGACA 2046
Db 300 GGGCCCTCGCTCATCTCTCTCTGACTGGTGGTCTCTGACGACTCATGTCTTCAGGATGACA 359
QY 2047 AAAATTTCAAGTACTCAGACTACAGTGTGGAGCGCTTCTCTGGGTCTCTGTCGACGAGA 2106
Db 360 AAAATTTCAAGTACTCAGACTACAGTGTGGAGCGCTTCTCTGGGTCTCTGTCGACGAGA 419
QY 2107 GCAAGCGCAGTGTCTCTGGGGTGCAGAGCTGAAGCTCAAACGCTATCATCACCCACCTT 2166
Db 420 GCAAGCGCAGTGTCTCTGGGGTGCAGAGCTGAAGCTCAAACGCTATCATCACCCACCTT 479
QY 2167 CTTCAATGATTTCACTTCGACTATGACATCGCTTGTCTGGAGCTGGAGAGTCTGCTG 2226
Db 480 CTTCAATGATTTCACTTCGACTATGACATCGCTTGTCTGGAGCTGGAGAGTCTGCTG 539
QY 2227 AGTACAGCACCGTCTGTGGCGCCCTCTGCTGCTGATGCTACCCATGTCTTCCTCTGCTG 2286
Db 540 AGTACAGCACCGTCTGTGGCGCCCTCTGCTGCTGATGCTACCCATGTCTTCCTCTGCTG 599
QY 2287 GC-AAGGCCCATCTGTGGTTCACAGCTGGGGGCACACAAAGAGGAGGT-ACCGAGCGCT 2344
Db 600 GCAAGGCCCATCTGTGGTTCACAGCTGGGGGCACACAAAGAGGAGGTCAACGAGCGCT 659
QY 2345 GATCCTGCAGAGGCTGAGATCGTGTCTATC-RACAGAGCACTGTGAGAGCTCATGC 2403
Db 660 GATCCTGCAGAGGCTGAGATCGTGTCTATCAACACGAGCGCACTGTGAGAGCTCATGC 719
QY 2404 CGCAGCAGATCACCCACGAAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACTCCT 2463
Db 720 CGCAG-AGATCACCCACGAAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACTCCT 778
QY 2464 GCGAGGTGATCTGTGGTGGCGCTTGTCAAGCGCGGAGAGATGGCGAATGTTCAGG 2523
Db 779 G-CAGGGTGAATCTGTGGTGGCGCTTGTCAAGCGCGGAGAGATGGCGAATGTTCAGG 837
RESULT 13
BF301347 1116 bp mRNA linear EST 21-NOV-2000
LOCUS 602029796F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4165103 5',
DEFINITION mRNA sequence.
ACCESSION BF301347
KEYWORDS EST.
SOURCE BF301347.1 GI:11247870
ORGANISM Mus musculus (house mouse)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9451 Row: 0 column: 24
High quality sequence stop: 686.
Location/Qualifiers
1. .1116
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4165103"
/lab host="DH10B (T1 phage-resistant)"
/clone lib="NCI CGAP SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
FEATURES
source
1. .1116
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4165103"
/lab host="DH10B (T1 phage-resistant)"
/clone lib="NCI CGAP SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:

```

Not1; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

```
BASE COUNT      325 a   315 c   302 g   174 t
ORIGIN
Query Match      23.7%; Score 735.8; DB 10; Length 1116;
Best Local Similarity 92.1%; Pred. No. 3.9e-134;
Matches 800; Conservative 0; Mismatches 62; Indels 7; Gaps 2;
QY 1137 CACTACCCGCCCAACATCAATGTCACATGCAATCAAGTGGTCCCAACCAACCGGAACGTTG 1196
Db 1 CACTACCCGCCCAACATCAATGTCACATGCAATCAAGTGGTCCCAACCAACCGGAACGTTG 60
QY 1197 AAGTGGCCCTTCAACTCTTCTATCTGTGGACCCCAACGACGATGAGTGGTCTCTGCACC 1256
Db 61 AAGTGGCCCTTCAACTCTTCTATCTGTGGACCCCAACGACGATGAGTGGTCTCTGCACC 120
QY 1257 AAGGACTATGTGGAGATCAACGGGGAGAGTACTCGGTGAGAGGTCCCAAGTTTGTGGTG 1316
Db 121 AAGGACTATGTGGAGATCAACGGGGAGAGTACTCGGTGAGAGGTCCCAAGTTTGTGGTG 180
QY 1317 AGCAGCAACAGCAGCAAGATTACATGTCACATTCATCTTCTGATCACTGTACAGGACAC 1376
Db 181 AGCAGCAACAGCAGCAAGATTACATGTCACATTCATCTTCTGATCACTGTACAGGACAC 240
QY 1377 GGGTTCCTAGCTGAGTACTCTCTACGACTCCACGACCCGTCGACGGATGTTCTATG 1436
Db 241 GGGTTCCTAGCTGAGTACTCTCTACGACTCCACGACCCGTCGACGGATGTTCTATG 300
QY 1437 TGCAGAGCTGAGCGGTGATCCGAAAGAACTGCGTGCAGCGGTGGGCGAGCTGCCCG 1496
Db 301 TGCAGAGCTGAGCGGTGATCCGAAAGAACTGCGTGCAGCGGTGGGCGAGCTGCCCG 360
QY 1497 GATTATAGTATGAGCGTTACTGCGCATGCAATGCCACCCACGATTCACGTGCAAAAAC 1556
Db 361 GATTATAGTATGAGCGTTACTGCGCATGCAATGCCACCCACGATTCACGTGCAAAAAC 420
QY 1557 CAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAAGCTGTGGGAGCGAAGT 1616
Db 421 CAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAAGCTGTGGGAGCGAAGT 480
QY 1617 GACGAGGAGGCTGACGCTGCTCTGCGAGTTTCAAGTGTTCGAATGGGAAGTGTCTC 1676
Db 481 GACGAGGAGGCTGACGCTGCTCTGCGAGTTTCAAGTGTTCGAATGGGAAGTGTCTC 540
QY 1677 CTTAGAGCCAGAGTGTAAATGGGAAGCACTGTGGAGATGGTCTGTGACGAGCTTCA 1736
Db 541 CTTAGAGCCAGAGTGTAAATGGGAAGCACTGTGGAGATGGTCTGTGACGAGCTTCA 600
QY 1737 TGTGACAGCGTGAATGTCTCTCTTGCACCAATATACCTACCGCTGCCAAATGGCCTC 1796
Db 601 TGTGACAGCGTGAATGTCTCTCTTGCACCAATATACCTACCGATGCCAAAGGCGCTC 660
QY 1797 TGTCTGAGCAAGGGCAACCCCTGAGTGTGTATGGGAAGACGGAATGTAGGATGGCTCCGAT 1856
Db 661 TGTCTGAGCAAGGGCAACCCCTGAGTGTGTAGGGGAAGACGGAATGTAGGATGGCTCCGCA 720
QY 1857 G---AGAAAACTGTGACTGTGGTGCATCTTTTACCAACAGGCTCGCGTGTGGT 1913
Db 721 GGACAAACACACTGGGACTGTGGGAGCGATCTTTTACCAACAGGATCAGGTGACAGGT 780
QY 1914 GGCACGAATGCG---GACGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCAGCCCTGGG 1969
Db 781 GGGACGAATTCGACAGCAAGGGAACAAAGTGGCCGTGCACGAGGCTCCAGGCTGTGTA 840
QY 1970 CCAGGCGCACTTGTGTGGGCGCTTCGCTCA 1998
Db 841 GAACAGGCCCAATGAGTGGGAACCCCTAGCA 869
RESULT 14
BI646380
```

```
LOCUS      BI646380      786 bp      mRNA      linear      EST 12-SEP-2001
DEFINITION 603276244F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5316539 5',
            mRNA sequence.
ACCESSION  BI646380
VERSION    BI646380.1 GI:15560616
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 786)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L14M11802 row: h column: 12
            High quality sequence stop: 784.
FEATURES             Location/Qualifiers
     source           1..786
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clones="IMAGE:5316539"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="VDH10B"
                     /clone_lib="NCI_CGAP Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
```

```
BASE COUNT      177 a   203 c   235 g   171 t
ORIGIN
Query Match      23.6%; Score 732; DB 12; Length 786;
Best Local Similarity 98.7%; Pred. No. 2.4e-132;
Matches 780; Conservative 0; Mismatches 5; Indels 5; Gaps 4;
QY 1409 CAACGACCCGTCGCCAGGATGTTCAATGTCAAGACTGACGGTGCATCCGAAAGAACT 1468
Db 1 CAACGACCCGTCGCCA-GGATGTTCAATGTCAAGACTGACGGTGCATCCGAAAGAACT 59
QY 1469 GCGCTGCGACGCTGGCGAGACTGCGCGATTATAGTATGATGAGCGTTACTGCGGATGCAA 1528
Db 60 GCGCTGCGACGCTGGCGAGACTGCGCGATTATAGTATGATGAGCGTTACTGCGGATGCAA 119
QY 1529 TGCCACCCACAGTTTCACTGTCAAAAACAGTTTCTGCAAGCCCTCTTCTGGGTCTGTGA 1588
Db 120 TGCCACCCACAGTTTCACTGTCAAAAACAGTTTCTGCAAGCCCTCTTCTGGGTCTGTGA 179
QY 1589 CAGTGTCAACGACTGTGGGACGGAAGTGACGAGGAGGCTGCAGCTGTCTCTGCTGGGAG 1648
Db 180 CAGTGTCAACGACTGTGGGACGGAAGTGACGAGGAGGCTGCAGCTGTCTCTGCTGGGAG 239
QY 1649 TTTCAGTGTTCCTCAATGGGAAGTGTCTCCCTCAGCCAGGAGTGTATGGGAAGCAAA 1708
Db 240 TTTCAGTGTTCCTCAATGGGAAGTGTCTCCCTCAGCCAGGAGTGTATGGGAAGCAAA 299
QY 1709 CTGTGGAGATGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCTCTTTGACCAA 1768
Db 300 CTGTGGAGATGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCTCTTTGACCAA 359
```

```

Qy 1769 ATATACCTACCGCTGCACAAATGCGCTCTGTCTGAGCAAGGCAACCTCGAGTGTGATGG 1828
Db 360 ATATACCTACCGCTGCACAAATGCGCTCTGTCTGAGCAAGGCAACCTCGAGTGTGATGG 419
Qy 1829 GAAGACGAGCTGTAGCGATGCTCCGATGAGAAAACTGTGACTGTGGCTGCGATCCTT 1888
Db 420 GAAGACGAGCTGTAGCGATGCTCCGATGAGAAAACTGTGACTGTGGCTGCGATCCTT 479
Qy 1889 TACCAACAGCGCTCGCTGGTGTGTGTCACAAATGCGGACGAGCGAGTGGCCCTCGCA 1948
Db 480 TACCAACAGCGCTCGCTGGTGTGTGTCACAAATGCGGACGAGCGAGTGGCCCTCGCA 539
Qy 1949 GGTGAGCCTCCAGCCCTCGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGA 2008
Db 540 GGTGAGCCTCCAGCCCTCGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGA 599
Qy 2009 CTGCGTGTCTCTGAGCTCATCTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTA 2068
Db 600 CTGCGTGTCTCTGAGCTCATCTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTA 659
Qy 2069 CAAG-ATGTGACGCGCTTCTGCGTCTGCTGCGACGAGCAAGCGAGTGCCTCTCGGG 2127
Db 660 CACGAATGTGACGCGCTTCTGCGTCTGCTGCGACGAGCAAGCGAGTGCCTCTCGGG 719
Qy 2128 TSCAGGAGCTGAAGCTCAAAAGTATCATCAACCCACCTTCTTCAATGANTTCACTTCG 2187
Db 720 TSCAGGAGCTGAAGCTCAAAAGTATCATCAACCCACCTTCTTCAATGANTTCACTTCG 776
Qy 2188 ACTATGACAT 2197
Db 777 ACTATGACAT 786

RESULT 15
BG915045
LOCUS 602814325F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4936590 5',
mRNA sequence.
ACCESSION BG915045
VERSION BG915045.1 GI:14295521
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: LLNL0869 row: a column: 07
High quality sequence stop: 793.
Location/Qualifiers
1..928
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4936590"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam4"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

```

```

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996).
BASE COUNT 216 a 258 c 273 g 181 t
ORIGIN
Query Match 23.2%; Score 721; DB 12; Length 928;
Best Local Similarity 93.2%; Pred. No. 3.3e-131;
Matches 832; Conservative 0; Mismatches 50; Indels 11; Gaps 7;
1886 CTTTACCAACAGGCTCGCTGTGTGTGTGCGACGAATGCGGACGAGGCGAGTGGCCCTG 1945
Db 1 CTTTACCAACAGGCTCGCTGTGTGTGTGCGACGAATGCGGACGAGGCGAGTGGCCCTG 60
1946 GCAGGTGAGGCTCCAGCCCTGGGCGGAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCC 2005
Db 61 GCAGGTGAGGCTCCAGCCCTGGGCGCA-GGCCCACTTGTGTGGGCGCTCGCTCATCTCTCC 119
2006 TGACTGGCTGTCTCTGAGCTCATCTGTCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 2065
Db 120 TGACTGGCTGTCTCTGAGCTCATCTGTCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 179
2066 CTACACGATGTGACGCGCTTCTGTGGTCTGTGTGACCAAGCGAGTGCCTCTGG 2125
Db 180 CTACACGATGTGACGCGCTTCTGTGGTCTGTGTGACCAAGCGAGTGCCTCTGG 239
2126 GGTGAGGAGCTGAAGCTCAAAAGTATCATCAACCCACCTTCTTCAATGANTTCACTT 2185
Db 240 GGTGAGGAGCTGAAGCTCAAAAGTATCATCAACCCACCTTCTTCAATGANTTCACTT 299
2186 CGACTATGACATCGCTTGTGTGAGCTGAGAAAGTCTGAGTACAGCAACCGTCTGTCG 2245
Db 300 CGACTATGACATCGCTTGTGTGAGCTGAGAAAGTCTGAGTACAGCAACCGTCTGTCG 359
2246 CCCCATCTGCGTCTGATGCTACCAATGCTTCCCTGTGTGGCAAGGCGATCTGGTTCAC 2305
Db 360 CCCCATCTGCGTCTGATGCTACCAATGCTTCCCTGTGTGGCAAGGCGATCTGGTTCAC 419
2306 AGGCTGGGGGCACAAAGAGGAGGAGTACCGGAGCGCTGATCTCTGCAAGAGGCTGAGAT 2365
Db 420 AGGCTGGGGGCACAAAGAGGAGGAGTACCGGAGCGCTGATCTCTGCAAGAGGCTGAGAT 479
2366 CCGTGTCTATCAACAGACCACTCTGTGAGGACCTCATGCCGAGAGATCAACCCACGAAT 2425
Db 480 CCGTGTCTATCAACAGACCACTCTGTGAGGACCTCATGCCGAGAGATCAACCCACGAAT 539
2426 GATGTGTGGGTTTCTCTAGTGGGGGTGTGGACTCTCTGCGAGGCTGACTCTGGT--GGC 2483
Db 540 GATGTGTGGGTTTCTCTAGTGGGGGTGTGGACTCTCTGCGAGGCTGACTCTGGTGGCC 599
2484 CCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGGAGTGGGGT 2543
Db 600 CCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGGAGTGGGGT 659
2544 GAAGGCTGCGTCTAGAGGAAACAG-CCAGGCGGTGTACA-CAAGGCTCTCTGTAGTGGG 2601
Db 660 GAAGGCTGCGTCTAGAGGAAACAGCCAGCGGCTGTACACCAAGGCTCTCTGTAGTGGG 719
2602 ACTGGATCAAGAGCACACTGGGGTATAGCAGATGGA-CAGACGCGGACCAACAAACAC 2660
Db 720 AACTGGATCAAGAGCACACTGGGGTATAGCAGATGGA-CAGACGCGGACCAACAAACAC 779
2661 CCACAGGATGCCCGACATGACACCTGGAT---ACAGGAGGAGAACTACGACATTT 2717
Db 780 CCACAGGATGCCCGAATGACACCTGGATTTACAGGAGGAGAACTACGACATCA 839
2718 ATGCTGTGGCTCTCCCCCCCCCAACACACACCTGTGAATCTGATCTCTTAG 2770
Db 840 TGCTTGTGGCATCCACCCCC---AAAAACAACAACCTGTGAATCTGATCTCTTAG 890

```


Search completed: February 21, 2004, 07:12:09
Job time : 6230 secs